1 *In defence of D-structure*

1.1 Introduction

While the relation between expression and meaning is fixed and arbitrary in words, in sentences it is mediated through syntax, a complex system. This then poses the question of how exactly that mediation takes place, whether it too is arbitrary despite the fact of being unfixed, or it is somehow sensitive to the particular derivational dynamics that result in phonetically expressible semantic concepts. Let's call this the Mapping Problem.

Successive stages of the generative enterprise have addressed the Mapping Problem differently. In the Standard Model, syntactic derivations were thought to start in a semantically articulated logical form, which after various adjustments reaches its phonetic form. In contrast, in the current Minimalist Program (MP), those two 'levels of representation' are situated as end-points of a forked derivation, which starts in a purely syntactic object constructed from a list of token words. The latter is the residue of yet a third conception, within the principles & parameters model, for which the starting point of the derivation was a full-fledged level of representation called D(eep)-structure. This level was taken to code lexico-conceptual relations, distinct from the scopal relations achieved through the derivation itself.

A priori, none of these answers to the Mapping Problem seem better or worse: it is an interesting empirical matter to determine which one wins – including possible alternatives. Minimalism, moreover, adds a further twist to the discussion: whichever answer works ought to be demonstrated as *an ideal* one, in some non-trivial sense, if the language faculty is to be seen in this conception as an optimal relation between meaning and its expression (sound or sign).

Wouldn't it be nice if optimality considerations alone could chose between alternative answers to the Mapping Problem? Unfortunately, we face a 'black-box' sort of puzzle. Whereas it is reasonably clear what we find in the expression side of things, what goes on in the Conceptual/Intentional domain(s) is harder to ascertain. Perhaps if we had as clear a picture on that side as we

1

have on the other, it would be easier to let sheer elegance decide on what the best possible mapping is between the two. But we have what we have, and part of our job is to deepen our understanding of the semantic side by sharpening our judgements about the syntax that carries us there – which we understand reasonably well.

This book can be seen as an exercise of this sort, which wagers for a particular answer: *semantics is 'opportunistic'*, a bit like a needy child. It won't drive the syntactic derivation (as in the Aspects Model) and it won't wait for it to finish either; rather, in a more dynamic fashion, it will take what it can literally as it can, for better and also for worse. This empirical thesis is actually not too far, at least reconstructing history a bit, from either Chomsky's original (1955) proposal – prior to the Aspects Model – or more accurately the cyclic systems that various researchers explored in the 1970s (Bresnan 1971; Jackendoff 1972; Lasnik 1972). In particular, this book will attempt to show that not only is what we can think of as 'higher-order' syntax (arising in context-sensitive terms as the derivation unfolds) subject to interpretation; simpler, 'lower-order' phrasal configurations are, too.[1]

In order to establish that 'dynamic' thesis, the book will start by analysing, in the present chapter, why the idea that there is a directional mapping from context-free D-structures to context-sensitive LFs was given up, or to what extent it really was. In section 2 of the chapter, the empirical reason for eliminating D-structure is presented, and next the technical notion of a 'level of representation' is analysed, and how that differs from a weaker notion of what may be thought of as a 'component' of grammar. Section 3 is devoted to providing arguments that the empirical discovery in point, while surely arguing against the notion of D-structure as a level of representation, has *no bearing on D-structure as a component of grammar*. In section 4 we can see how, despite his rhetoric within the MP, Chomsky himself effectively assumes all the defining characteristics of a component of D-structure. Section 5 examines the question of whether relevant arguments for complex derivations involving conditions on their input remain solid within the program. Section 6 considers different possible ways of coding D-structure information: in LF, in the lexicon, in a separate component of D-structure. Conclusions are presented in section 7, where it is suggested that either we should eliminate D-structure residues within the MP, or else explore ways of rationally incorporating them.

1 By 'context' we mean, here, the derivational history of a given phrase-marker. Context-free relations pay no attention to such nuances, while context-sensitive ones (for instance transformations) must.

The rest of the book can be seen as a long argument for that second position. The alternative one is certainly coherent, and equally minimalist,[2] but it should provide ways to explain away the evidence adduced here in favor of the more dynamic view. At any rate, these issues must be raised because, if semantic interpretation is to be dynamic and opportunistic, it won't wait until the end of the derivation. But then we must show that it is profitable for semantics to start interpreting stuff early on, and moreover that, if the system is minimalist, it does so for good reason.

1.2   Why D-structure was eliminated

Let's start with a bit of history, to put things in perspective.

1.2.1   Kearney's example and Lasnik's argument

An unpublished example by Kevin Kearney dramatically questioned the level of D-structure, as understood in the principles and parameters (P&P) system. The context for the argument was Chomsky's (1981, p. 309ff.) discussion of 'Tough'-movement:

(1)   John is easy (for anyone) to please.
       (cf. It is easy (for anyone) to please John.)

Chomsky reasons that:

> the matrix [*John*] is not inserted at D-structure, but is also not moved to the matrix subject position. The only resolution to this paradox . . . is to assume that lexical insertion of the matrix subject is at S-structure in this case . . . We are therefore led to the conclusion that lexical insertion can take place freely. [p. 313]

This is correctly considered a way to 'simplify the theory of lexical insertion'. That said, however, observe Kearney's example:

(2)   A man who is easy (for anyone) to please is easy (for anyone) to convince.

One might be persuaded to insert *John* anywhere in the derivation, but it is harder to accept the *lexical* insertion of *a man who is easy (for anyone) to please*, in the very same context where *John* creates a problem in (1). Howard Lasnik raised this difficulty with Chomsky through personal communication, and they both realized the need for traditional *generalized transformations*, of the sort postulated in Chomsky (1955), to handle the difficulty. If the latter

2 See in particular Hornstein *et al*. (2005), and its references.

are part of the grammar, we can (in current terms) merge not just *John*, but also a phrase of arbitrary complexity, such as the (previously assembled) *a man who is easy (for anyone) to please*. But conceding that point undermines D-structure, for the whole purpose of that level of representation was to create a *unified* object to express configurational relations – crucially prior to the occurrence of transformations.

D-structure's troubles started much earlier, however, when generative semanticists proposed getting rid of it.[3] Their classic argument went as follows. Assume syntactic principle P (it does not matter which). Show how assuming P explains why an unattested word *glarf* – whose meaning appears to be composed of X plus Y – may actually *not* arise by relating X to Y, as this would violate P (concrete arguments on this are reviewed in section 2.2.2). Therefore objects like *glarf* – that is, words – do not matter for the linguistic system as much as *the component elements* X and Y, moreover as related through principles like P in relevantly viable conditions. If so D-structure (a collection of grammatical counterparts to *glarf* arranged in terms of principles like P) is a spurious generalization. The linguist must, rather, study X, Y and other similar 'thoughts', and how they arrange themselves to compose surface language.

The fate of that proposal is well-known (see chapter 2), although it is not often admitted that much of the aftermath gave way to assumptions and conditions that are, on the surface at least, not unlike the rejected ones.[4] Differences, of course, exist; for instance, meaning is nowadays typically read off output syntax (LF), whereas in those days it was coded in the input – and that, we will see shortly, makes a significant difference. That said, this book will try to show that the overall problem that led to generative semantics is far from resolved.

1.2.2   Levels and components

The fact that we cannot maintain a level of D-structure, for empirical reasons, does not entail that we should get rid of D-structure *information*. That issue would arise anywhere else in the grammar, too. Suppose some crucial empirical argument A questions LF as a level of representation. That in itself does not argue for the need to eliminate an LF *component*, mapping (say) S-structure (or

---

3 See for instance McCawley (1968), Lakoff (1971) or Postal (1972), and also Newmeyer (1980, 1996), Harris (1993), Huck and Goldsmith (1995), and Barsky (1997) for related discussion.

4 This point is critically raised in Seuren (2004); for a more sympathetic view of minimalism, though still admitting connections to generative semantics, see Lasnik and Uriagereka (2005).

any other input component) to the semantic interface. It is entirely possible that A may have nothing to say about the LF component proper, and instead it only affects LF strictly as a unified level of representation.

To see this in detail, let us remind ourselves of what a *level of representation* is, technically, and compare it to a less strict component of the grammar. That way we may be able to decide whether Lasnik's point with Kearney's example does argue, specifically, against the level of D-structure, or it has instead a more radical, and eliminative, consequence on D-structure information at large within the system.

The way Chomsky speaks of those notions is currently tentative: 'A standard assumption,' he observes, 'is that UG [Universal Grammar] specifies certain *linguistic levels*, each a symbolic system ... Each linguistic level provides the means for presenting certain systematic information about linguistic expressions. [UG] must specify the interface levels ... the elements that constitute these levels, and the computations by which they are constructed' (Chomsky 1995, p. 167ff.). Two things are noteworthy here. First, that Chomsky speaks of a 'standard assumption', thus something that may or may not be acceptable as the MP unfolds – with its emphasis in questioning, precisely, traditional, often unmotivated, assumptions. Second, the paragraph just mentioned would read very similarly if we substituted the word 'component' where it uses 'level'. That is, if all we are talking about is the *systematicity* of a given sort of information in a linguistic expression, and how it is organized as it relates to other systems, then we are not really speaking of the technical notion *level of representation*. The more unassuming notion *component* would do.

In *Aspects of the theory of syntax* (1965, chapter 3, fn. 2), Chomsky sums up what we commit to when we seriously speak of levels of representation:

> Linguistic theory provides a (universal) system of *levels of representation*. Each level L is a system based on [1] a set of primes (minimal elements – i.e., an alphabet); [2] the operation of concatenation, which forms strings of primes of arbitrary finite length (the terms and notions all being borrowed from the theory of concatenation algebras – cf. e.g., Rosenbloom [1950]); [3] various relations; [4] a designated class of strings (or sets of strings) of primes called L-markers; [5] a mapping of L-markers onto L'-markers, where L' is the next 'lower' level (thus levels are arranged in a hierarchy). In particular, on the level P of phrase structure and the level T of transformations we have P-markers and T-markers ... A hierarchy of linguistic levels (phonetic, phonological, word, morphological, phrase structure, transformational structure) can be developed within a uniform framework in this way. For details, see Chomsky (1955). [*Numbers added.*]

6   *Syntactic Anchors*

Note, first of all, that all of that talk is within *concatenation* algebras. Strictly, none of these notions survive untouched as we move towards what may be called 'merge algebras', which haven't been systematically formalized yet (see chapter 7 for more detailed discussion). Being generous with the formalism, though, we can translate the relevant notions, and much of the paragraph above – specifically an adapted notion 'level of representation' – will survive. Concretely, [2] is where adjustments need to be made, given Minimalist 'bare phrase-structures', which do not involve concatenation (in fact the system takes linear order to obtain only in the PF string); nonetheless, one could relax the definition so as not to demand a linearly organized set of categories to operate on, resulting from Merge.[5] That qualification made, [1] and [3] are straightforward – for levels or for more unassuming components. All that [1] and [3] say is that these are substantive parts of the system, in the primitives they use (symbols from some alphabet) and the relations they allow (e.g. phrases). Whether [5] obtains *trivially* (with no mapping) is an architectural decision; standard minimalism insists on the idea that substantive parts of the model must interface with *language external* systems, and if so the ultimate need for [5] depends on the empirical details of those. If the model is organized as in Chomsky (1995), [5] is irrelevant; but even if [5] did hold, it could be expressed equally well for levels or for less articulated components (e.g. a chain, by definition, is structured from a phrase, whether or not either one of those – phrase or chain – is integrated into a level of representation). Finally we are left with [4]. This is really it: what distinguishes a level from a mere component is whether we can justify *a designated class* of (sets of) strings. There is a *unification* implied in the characterization of that designated class of objects; D-structure is a designated class of phrases; LF is a designated class of chains; PF is a designated class of speech objects and so on. There is no meaningfully unified class of D-structures (P-markers) or S-structures (T-markers) in the derivation of any given sentence. Those *collections* of classes can be regarded as a (weaker) *component* of the system, but not one of its levels of representation.

---

5 Concatenation is normally meant to correspond to mere 'derivational activation' in a one-step-at-a-time (Markovian) derivation, yielding sets of strings. The relevant ensuing linearity need not translate into PF linearity, although it can (and if so it would arguably run in the opposite direction to Kayne's (1994) Linear Correspondence Axiom; see chapter 6 on this, which can actually be used to address a puzzle that arises with adjunct linearization). In the revised minimalist system, derivational activation should, instead, correspond to merged set-theoretic objects.

Although a formal definition is probably unnecessary at this point, for concreteness we may understand a component of the system, in effect, as any kind of information that is syntactically analysable *as a coherent unit*. A level of representation is, of course, that and more, as we have seen: essentially the idea that all such components in a given derivation constitute a unified whole of some sort, from which further operations may be defined. Apart from asking which is a more empirically adequate tool to model the workings of the language faculty – a level or a component – the minimalist must also ask which, if any, is a priori more desirable. If simplicity is relevant, then surely the component trumps the level: its characterization makes fewer commitments and the resulting objects are less complex, in computational terms (a component does not have to involve a collection of objects of *in-principle-arbitrary complexity*).[6] Therefore, one must find strong empirical arguments in favor of the notion of a level of representation if it is to be preferred over that of a corresponding system component.

1.3   Rethinking the case against D-structure

The Kearney/Lasnik argument against D-structure is very serious: 'Tough'-movement shows that we cannot have a 'unified' notion of D-structure. However, as we have just seen, that is in fact the central difference between a level and a component. That said, we must ask a related question.

1.3.1   Is Kearney's an argument against a D-structure component?
Note that each of the sub-parts one has to construct in order to undergo the generalized transformation that Lasnik proposed as a solution to Kearney's puzzle was, in itself, a sort of 'micro-D-structure', prior to the application of transformations. To put it differently, in that pristine derivational moment when only Merge has occurred, *whatever we have* is arguably a D-structure analogue. Granted, that is not a D-structure *level*, but it may perfectly well be a D-structure 'chunk' of the system – part of a D-structure component.

Whether that conclusion ought to be maintained depends on whether, *in the input to derivations*, there is reason to believe that a part of the language faculty is responsible for coding a system based on [1] a set of primes and organized in terms of [2] some associative operation that forms collections of primes of

---

6 An immediate issue arises in what may be thought of as an 'address' problem: how the various chunks of structure that arguably spell-out separately are 'put together' into a coherent unit. See chapter 4 on this.

arbitrary finite length such that [3] non-trivial relations among them can be established. The answer to all these questions is empirical. Yet Chomsky has maintained, in his minimalist works, that there is a *conceptual* reason to answer them prior to investigation. By assumption the language faculty arises only as an optimal device connecting points of interface with extra-linguistic systems. If there is *no point of interface* between the syntactic engine and an outside system that would code information relevant to D-structure, then there should be no such component. On the other hand, we could challenge the major premise. Could there be an interface that is relevant to our purposes? If the answer were Yes, it would be legitimate to postulate a D-structure to represent that juncture.

That option is occasionally forgotten. As noted above, minimalism explores the optimal connection between sound and meaning. Of course 'sound' and 'meaning' interfaces are *empirically* established. In that sense, if it turned out that, in fact, the system also interfaces, say, a visual, tactile, or musical capacity, for all the linguist should care, then so long as the interface and what drives the system there are both optimal, the minimalist should be satisfied. In other words, the key is not in relevant interfaces being 'sound' and 'meaning', specifically; rather, *whatever the interfaces happen to be*, the minimalist issue is accessing them in an optimal way. The empirical task, needless to say, is then to determine what the actual interfaces are.

1.3.2   What are the linguistic interfaces?

In short, nobody knows. That is already important, since although the question is perfectly empirical, the answer should not be a truism, when so much is at stake. These matters are familiar on the 'sound' side of the grammar. When we say that there is a *single* level of PF that interfaces with both the perception and articulation of phonetic sound, what we are subscribing to, in essence, is the Analysis by Synthesis theory (see Poeppel *et al*. [in press] for a recent perspective and references). This is an important empirical finding, which researchers naturally argue for and against. Note in this respect that much important work on signed languages, whose communicative modality is – at least on the surface – quite different, hasn't immediately entailed a crisis on the 'externalization' side of grammar. The relevant interface is or is not of the same sort, depending on whether it targets speech or manual signs; the matter is interesting regardless of what the answer is, and totally consistent with the MP if the relevant mapping is optimal in some appropriate sense. And if it turned out that language could also be externalized on a tactile system, say, so long as we are still within optimality parameters, absolutely nothing changes in the overall reasoning.

Curiously, on the 'meaning' side of the grammar we often do not even pose these questions. Of course, nobody has a complete enough picture of what is implicated in meaning to seriously entertain the possibility that it is coded in terms of one, two, or seven components, alternative or universal. There is, no doubt, a vague intuition that meaning deals with 'intentional' information (of the sort involved in reference, the calculation of truth, and so on), and 'conceptual' information (which distinguishes predicates from arguments, and, among those, sorts of either, and so forth). Surely there are very elaborate formal systems that pack these notions into mathematical artifacts. Whether that does have much, little, or nothing to do with the way human beings actually 'mean' things is anybody's guess. To jump from that to the conclusion that 'there is a *single level of representation* that the system has to optimally interface' seems hasty.

In part the hastiness may come from a traditional philosophical perspective that was never central to the program of structural linguistics, and becomes even less so within the MP. This is the idea that knowledge of language may have something to do with linguistic representation. Philosophically, a representation can be seen as a relation between a subject and a theory of a (formal) language; in that sense, representation is *by definition* intentional, the sort of relation existing between a 'symbol' and 'what it stands for'. But there are two reasons why all of this may not be relevant to the minimalist characterization of the language faculty. First, not even purely linguistic notions ('feature', 'category', 'phrase', 'c-command', and so on) stand in any 'symbolic' relation with regards to what their import is within the grammar; there is no obvious, useful, linguistic gain in saying that 'the feature [+consonantal] represents such-and-such'. Surely an important naturalistic issue arises as to what [+consonantal] (or any such notion) ultimately *is*, but calling its nature representational (in the philosophical sense) hasn't had any consequences. Second, when it comes to more familiar 'reference'-bearing elements, or more generally elements with a denotation (*John*, *book*, *sing*, and the like) not a single theory exists that clarifies what sort of relation takes place between such words and what they denote. This may seem too harsh a comment to make, but the truth is all familiar theories – all of them, really – *presuppose* the relation, in statements of the sort of *'snow' refers to snow*. Indeed, contemporary semantics prides itself in being able to operate without detailed knowledge of that intentional dependency, solely on the basis of the algebraic combinations that work above those relations, which are deemed either irrelevant – where the theory bottoms out – or something that psychologists ought to work out the details of. But the details themselves are not forthcoming.

At any rate, one can call *those details* representational too, and in that sense assume their intentionality again by definition; this however will not tell us even the most basic properties of how the referential process is achieved. For instance, it is customarily assumed that *book* merely denotes the set of books, and in that sense denotation is in large part presupposed as a lexical process, to be turned into genuine reference the moment some determiner is compositionally added to one such predicate. Thus *a book* is supposed to be able to refer to whatever is in the reader's hands, in essence quite independently of whether the reader is actually using that expression in any given derivation and any given situation.

That may, in the end, be the way human intentionality works; then again it may not. It could actually be the case, for instance, that a variety of semantic cues are built into the, as it were, 'ultimate reference' of an expression: not just purely conceptual issues like its being a book, but also more contextual ones, like its being conceptualized in a given speech act qua its physical characteristics (as opposed to intellectual ones); or even duller ones, like whether the speaker has decided to place the phrase intending to refer to the book in a prominent grammatical region, like the clausal left-periphery. If all of that does contribute to the phenomenon, intentionality may well be a *process*, rather than a pre-established (lexical) category. As is customary in science, this has to be established empirically, not assumed as dogma handed down by tradition, venerable as it may be.

It will take us some time to get to the level of certainty with which questions like these are asked on the sound – or more accurately 'externalization' – side of grammar, if only because the answer is (probably) buried deeper inside our mind/brains, and at any rate linguists have been focussing on it for less time. Nonetheless, there are some obvious considerations one can already raise. As mentioned above, generative studies once suggested a model that usefully separates broadly semantic notions in terms of, on the one hand, purely conceptual information (D-structure) from, on the other, the ultimate intentional information that it leads to (LF), and in fact specifically maps the latter from the former (from the Extended Standard Theory on). That conception is either right or wrong, but certainly not senseless.

Suppose we had reason to believe that a system we are studying – say, vision – has components to differentiate the perception of color, depth, outline, and so on. Then surely two possible, extreme, ways to proceed suggest themselves. A 'modular' approach would try to relate *each* of the observed components, through some 'deep' array of relations; a 'holistic' approach, in contrast, would try to pack all the information together. Neither of these is