DRAFT OF 17 OCTOBER 2002

Some Consequences of Compositionality

James Higginbotham

University of Southern California*

1. General Discussion.

According to the view that I advance here, the thesis that the semantics of human first languages is compositional should be understood as imposing a *locality* requirement on the computation of semantic values. So understood, it may form part of a restrictive theory of semantics, intended to contribute to an account of human linguistic competence, and first-language acquisition. Taken to extremes, however, compositionality is apparently false, both for humdrum examples such as those given below, and for more tendentious cases such as those suggested in Higginbotham (1986), and others, only a couple of which I shall be able to consider here. Moreover, there are significant problems even in formulating the compositionality thesis for long-distance linguistic relations. My aim here will not be to examine all of these points in detail, but rather to clarify what I take to be at empirical issue in the discussion.

The (by now considerable) literature on compositionality does not tire of repeating that compositionality holds if "the interpretation of a sentence (or discourse) is determined by the meanings of its parts, and the way they are combined," or other words to that effect. Nor does it tire of stating that, so understood, the compositionality thesis verges on the trivial. For, there being nothing else but the parts of a sentence and the way they are combined to *give* its interpretation, what else could interpretation be determined by? The thesis is not quite trivial, however, inasmuch as it presupposes that there is such a thing as "the interpretation of a sentence or discourse" in a proper, context-independent, sense. Thus, one could propose a point of view according to which semantics, or what is determined independently of context in virtue of linguistic form, and pragmatics, or what is supplied within context, are treated together in an overall account of human communication; and in that case it could be that the semantic part of such a theory fails of compositionality, although the theory as a whole is compositional with respect to the structure of communicative acts in context. I take it that some applications, for instance in Discourse Representation Theory, may actually have this property. At any rate, I shall concentrate here on semantics proper, assuming that there is an autonomous theory of those aspects of interpretation that are determined by linguistic form, and that the weak sense of compositionality quoted above therefore goes without saying.

It is customary, and correct, to distinguish between *lexical* semantics, the account of the meanings of whatever the most primitive elements of language are, and *combinatorial* semantics, the account of the projection of meaning from constituents to complex structures. Some care must be taken, because what exactly constitutes "the lexicon" is a theoretical matter; however, I will assume in what follows that the interpretations of lexical items are either given outright, or else constructed in some way for which the question of compositionality does not arise.

Suppose then that semantic interpretation, or more precisely those aspects of interpretation that are strictly determined by linguistic form, takes for its input standard structures, trees with labels on the points, and with certain relations between the points. Let **T** be such a structure. What is it for the semantics of **T** to be "compositional," in something other than the trivial sense? And what is it for the semantics of a whole language to be compositional?

I will assume that *local compositionality* is a property of grammars **G**, effectively stating that there is a function $f_{\mathbf{G}}$ such that for every syntactic structure **T** licensed by **G**, and every constituent X=Y-Z in **T**, where Y and Z themselves are possibly complex, and X is possibly embedded in larger structures W=...X..., up to **T** itself, the family $M(X)$ of meanings borne by X in **T** is the result of applying $f_{\mathbf{G}}$ to the ordered triple $(F,M(Y),M(Z))$, where F comprises the formal features associated severally with X, Y, and Z, and $M(X)$ and $M(Y)$ are the families of meanings borne by X and Y in **T**.

If the meanings of lexical items are invariant, and the formal features of no complex X are also those of a lexical item, then the qualifier "in **T**" can be dropped. I will assume both of these propositions. The invariance of a lexical item is, one would expect, part of what is intended by calling it *one* lexical item in the first place; and the formal features of lexical items can always be distinguished, say by marking them and no others as "+lexical."

We can put further conditions on compositionality by supposing that meaning is *deterministic*; i.e., that for each point X in a tree **T**, the meaning of X is unique. The assumption of determinism has immediate syntactic consequences, inasmuch as ambiguities of scope will have to be syntactically represented. Moreover, their representation cannot take any simple form: at least for liberal speakers like me, it is pretty easy to construct, for any n, a sentence having scope-bearing elements $A_1, ..., A_n$ such that, if we go by apparent surface constituency, A_j is within the scope of A_i iff $i<j$; but from the point of view of interpretation A_j is within the scope of A_i iff $j<i$; and it

follows that no finite set of formal features attached to the A_i can replicate the intended relation of scope-inclusion. (Of course, intuitions give out as n gets bigger: but even simple examples with n=4, such as *Everybody didn't answer many questions about each book*, meaning that for each book, there are many questions about it that not everybody answered, suffice to make the point.) I will assume determinism in what follows, and, I hope without loss of generality, that relative scope is determined in the standard way, by c-command at LF.

Thus far I have spoken fast and loose about the meaning or range of meanings $M(X)$ of a point X. I now want to spell out further what I mean by this. I assume, as in other work, that what is crucial is not "meanings," whatever they may be, but rather what it is to know the meaning of an expression. As is customary, even if surely an idealization, I will assume that knowledge of the meaning of an expression takes the form of knowledge of a certain condition on its reference (possibly given various parameters) that is uniform across speakers; that is, that we can sum up what is required to know the meaning of an expression in a *single* statement, and that we do not adjust that statement so as to reflect, for instance, the different demands that may be placed upon speakers, depending upon their age, profession, or duties.

The form that conditions on reference will take will vary depending upon the referential categories of lexical items, and the categories of X, Y, and Z in a constituent X=Y-Z. Where Y with formal features $F(Y)$ is the root of $\mathbf{T_1}$, and Z with formal features $F(Z)$ is the root of $\mathbf{T_2}$, let $\mathbf{T}$ result by combining, or "merging," these, so that $\mathbf{T}$ has root X with formal features $F(X)$, and for immediate successors just Y and Z. We put $\mathbf{T}$=$\mathbf{T_1}$-$\mathbf{T_2}$, noting in particular that the formal features of *Y* and *Z* are unchanged following the merger.

For an example (ignoring tense): suppose *X*=*John*, *Y*=*walks*, and our tree **T**, in relevant detail, looks like this:

```
      X, <1>->Y, <>
       / \
      /   \
     Y    Z, <1>
     |      |
   John   walks
```

where the sole formal feature of *Z* is the thematic grid represented by ‘<1>’, announcing that *Z* is a one-place predicate, and the sole formal features of *X* are the indication of discharge of the θ-position <1> by the θ-marking of *Y* by *Z*, and the empty thematic grid <> appropriate for a closed sentence. We may then rehearse the familiar litany of referential semantics, the crucial combinatorial statement being that *X* (strictly, the tree with root *X*) is true just in case the reference of *Y* satisfies *Z*.

Thus, in a semantics that purports to characterize, in a finite theory, what we actually know about language, the elements that were called the meanings of constituents above are replaced by statements giving what someone knows who knows the meanings of those constituents.

2. Comparisons, Extensions, and Examples.

A number of authors, including those of work cited below, have latched on to the idea that compositionality should require (in the terminology adopted here) that whenever the meaning of the root *Y* of $\mathbf{T}_1$ is the same as that of the root *Y’* of $\mathbf{T'}_1$, the meaning of the root *X’* of $\mathbf{T'}=\mathbf{T'}_1\text{-}\mathbf{T}_2$ should be the same as that of the root *X* of $\mathbf{T}=\mathbf{T}_1\text{-}\mathbf{T}_2$; likewise for *Z*.

Local compositionality in the sense defined above, however, does not require this: the reason is that it is perfectly possible for *Y* and *Y'* to have the same meaning, but different formal features; and in this case, either for that reason alone or because the roots *X* and *X'* receive different formal features in consequence, meaning may not be preserved. I think this point is worth taking notice of, because otherwise one will face strange counterexamples. For instance, I suppose it will be agreed that the simple N *autobiography* is synonymous with the complex N *history of the life of its own author*. Even so, if I read John's autobiography, I do not read John's history of the life of its own author, and indeed the expression *John's history of the life of its own author* is both ungrammatical and meaningless (for want of an antecedent for the anaphor *its own*). The basis for the inequivalence is, of course, that whereas *history of the life of its own author* is a one-place predicate, where the antecedent of *its own* is the open position in the head N *history* and ultimately the element θ-marked by that N, as in *That book is a history of the life of its own author*, the N *autobiography* is descended from the inherently relational N *biography*, so that *John's biography* may be interpreted as referring to a history of the life of John, and *John's autobiography* as referring to a history of the life of John written by John.

Local compositionality, as I have defined it, is not a property of human languages in general. The simplest counterexample known to me is (1):

(1) John may not leave

There is no serious question but what the constituent structure of the predicate is as in (2):

(2) [may [not leave]]

Nevertheless, the sentence, understood deontically, may not be uttered as giving John permission not to leave, but can only be meant as denying him permission to leave. By local compositionality, the constituent *not leave* must amount to the negation of *leave* (for that is how it works in metaphysical examples, *John must not have left*, for instance). But then, if deontic *may* is understood as it would be in *John may leave*, the wrong interpretation is generated.

Our example is of a sort not confined to English (similar examples with deontic verbs are found in Italian, for instance), but it is easy to get around it all the same: just cut the chunks for compositionality a little more coarsely, or give negation different interpretations, only one of which fits under the modal. Various expedients are possible. In fact, these expedients are characteristic of some of the literature that employs higher-order logic whilst intending to have the syntactic inputs to interpretation close to the surface. But there is a larger moral to the story.

In formulating a restrictive semantic theory, we face a problem not in one or two unknowns, but three. There is, first of all, the problem of saying precisely what the meaning of a sentence or other expression actually is, or more exactly what must be known by the native speaker who is said to know the meaning. Second, there is the question what the inputs to semantic interpretation are: single structures at LF, or possibly complexes consisting of LF structures and others, and so forth. And third, there is the question what the nature of the mapping from the inputs, whatever they are, to the interpretation, whatever it may turn out to be. In the inquiry, therefore, we must be prepared for implications for any one of these questions deriving from answers to the others.

The model for some recent discussion, taken from formal language theory and indebted to the work of Richard Montague, seems to me deficient in respect of the empirical problems here. Thus Hendricks (2001), a recent discussion carrying on that tradition, commences from a point of view that takes the syntactic inputs to be given by a term algebra over some alphabet; the interpretation of expressions not generated by the algebra is not considered. In a similar vein, Hodges (2001) adopts the axiom (p. 11) that only *grammatical* expressions are meaningful, an axiom that, together with the thesis that synonymous expressions are intersubstitutable, leads him to deny, by deduction so to speak, that apparently synonymous expressions really are synonymous. In support, he gives the quartet (3)-(6) (taken from Gazdar, Klein, Pullum, and Sag (1985)):

(3) It is likely that Alex will leave.

(4) It is probable that Alex will leave.

(5) Alex is likely to leave.

(6) *Alex is probable to leave.

which leads to the conclusion that *likely* and *probable* are not synonymous. This, I think, is a mistake: *likely* and *probable* are synonymous (with some minor qualifications), and all that is happening is that *probable* does not admit subject-to-subject raising. So (6), ungrammatical as it is, is perfectly meaningful, and in fact synonymous with (5). (In Higginbotham (1985) I argued explicitly that semantics is indifferent to questions of grammaticality; but nowadays this is often assumed in practice anyway.)

Another example from Hodges, where he argues to the same effect as through (3)-(6) above, and reaches what I believe to be the right conclusion, but for the wrong reasons. Again the data are from Gazdar *et al.* (1985):

(7) The beast ate the meat.

(8) The beast devoured the meat.

(9) The beast ate.

(10) *The beast devoured.

Since *eat*, but not *devour*, undergoes object deletion, Hodges concludes that what he calls Tarski's principle, which would imply in view of the interchangeability of *ate* and *devoured* in the context *the beast ... the meat*, that they should also be interchangeable in the context common to (9) and (10), is false. The choice of the equivalent or near-equivalent verbs *eat* and *devour* is not accidental: Hodges goes on to suggest the weaker condition that we might aspire to what he calls a *Husserlian* semantics, which he defines (I simplify somewhat) as one in which synonymous expressions have the same semantical categories. The conclusion then is that, despite first appearances, *eat* and *devour* are not synonymous.

Now, it happens that Hodges' conclusion is at least partly correct. The verbs are not synonymous, at least if they are both considered single words, because *devour* is inherently telic, an accomplishment verb in Vendler's terminology, whereas *eat* is not. For this reason we have *the beast ate at the meat*. but not **the beast devoured at the meat*, we have *John ate up the applesauce*, but not **John devoured up the applesauce*, etc. The ambiguous behavior of *eat* is a matter for discussion, with Krifka (1992), for instance, arguing for a single, underspecified, lexical entry, whereas I myself in Higginbotham (2000) took the view that the V *eat* itself was ambiguous. Supposing that we adopt the latter view, there is still an issue in the distinction between *eat* and *devour*, because (9) admits a telic interpretation (as shown, e.g., by *John ate in ten minutes*), and the V of (7) may of course be telic. Stripping away as irrelevant to our purposes the haze of usage or coloration that distinguishes the verbs, as for instance that *devour* is a learned word, whereas *eat* is not, we are left with something of a defensible synonymy between *devour* and telic *eat*, and a grammatical distinction between them that by Hodges' lights undermines the Tarski principle. Because he intends to support the weaker Husserlian principle, Hodges concludes that the V must be declared non-synonymous. This conclusion is in fact forced upon him by the axiom that only grammatical expressions are meaningful. The truth about these cases, however, is more complex.

A further semantic principle that has been suggested, for instance in the textbook Heim and Kratzer (1998), is that the modes of semantic combination should be restricted to just function and argument. In their setting, however, which assumes a logic of order ω, this principle, for a language that is locally compositional, can always be satisfied. For, suppose that X=Y-Z, and that $M(X)=f_{\mathbf{G}}(F,M(Y),M(Z))$, as above. Then the formal features F together with $f_{\mathbf{G}}$ give us some function, f_{127} say, so that where $M(Y)=P$ and $M(Z)=Q$, $M(X)=f_{127}(P,Q)$. We may then define a function P^* by: $P^*(Q)=f_{127}(P,Q)$, and set $M(Y)=P^*$ instead of P. (Carried out recursively, this will mean adjusting interpretations all the way down to the leaves of the tree with root Y.) The proposed principle therefore doesn't add anything.

The construction of P^* depends upon the availability of functions of higher order, so that if these are eschewed then the principle acquires force. Then, however, it is false, even for trivial examples such as adjectival modification: in an expression such as *black cat*, neither element is an argument of the other, both being simple predicates. It is also false for other cases where what Otto Jespersen called the *nexus* between constituents is inexplicit, as it is for instance in explanatory, purposive, resultative, or possessive contexts, illustrated in (11)-(14).

(11) Having unusually long arms, John can touch the ceiling [explanatory nexus]

(12) I bought bones [*O* [PRO to give *t* to the dog]] [purposive nexus]

(13) John ran himself ragged [resultative nexus]

(14) John's boat [possessive nexus]

The question of nexus apart, my own preference is for a system with a small family of restricted choices of semantic combination, and a weak second-order logic (one in which quantification over predicates and property-abstraction are possible, but predicates do not occupy argument positions). The question will then arise whether the combinatorial system for human languages is universal, and so fixed in advance (which is, at least in practice, the standard working hypothesis).

The choices of semantic combination certainly include the following, for structures X=Y-Z, as in Higginbotham (1985):

(θ-marking) *Z* is an argument of *Y*, so that where *Y* expresses the condition on reference ...*y*..., *X* expresses the condition ...*M*(*Z*)..., where *M*(*Z*) is the reference of *Z*.

(θ-identification) Some open position in *Z* is identified with an open position in *Y*, so that, where *Z* expresses some condition ...*z*..., and *Y* some condition ___*y*___, *X* expresses the conjunction ...*y*... & ___*y*___.

(θ-binding) *Z* expresses some condition ...*z*..., and *Y* binds the position marked by '*z*', so that, where *Y* expresses the condition ___*A*___, *A* a predicate position, *X* expresses the condition ___λz (...*z*...)___.

All of these operations may be generalized in various ways (e.g., we can allow several instances of θ-identification at once, or identification together with θ-marking, etc.). However, all are to take place under sisterhood, so that local compositionality is preserved. In my (1985) I added the (somewhat unfortunately named) case of *autonymous* θ-marking, the idea being that *X* would express the condition, not ...*M*(*Z*)... but rather ...**T**..., where **T** is itself the tree with root *Z*. This move is unnecessary, however, if we have set down that *M*(*Z*)=**T** in the first place. I thought then (and think now, though I shall not defend this view here) that complement clauses, of the type to which the account was to apply, referred to themselves; though we must add the stipulation, or convention, that they are to be understood as if uttered.

It is the possibility of θ-marking, as opposed to the possessive nexus, that distinguishes *autobiography* from *history of the life of its own author*. The latter is effectively construed as in (15):

(15) (the *x*) history of the life of its own author(*x*) & **R**(John,*x*)

where **R** can be anything, whereas *John's autobiography* is:

(16) (the *x*) autobiography(-of) (John,*x*)

The distinction between the two is already expressed in their formal features: *autobiography* has two open positions, *history of the life of its own author* only one. At the same time, these expressions are synonymous, considered as predicate nominals.

Thus far, I have spoken of constituents *X*=*Y*-*Z* independently of their syntactic categories, and independently in particular of which, if any, is the head of the construction. We may, however, consider the notion of *headedness* as it might be applied to the semantic domain, as well as the syntactic. I suggest that it is possible for the semantic head of a construction to be other than its syntactic head, and that a number of the arguments that have been made in favor of higher types in semantics are answered by noting that the exceptional cases that seemed to call for raising to higher types are in fact merely cases of switch-headedness; i.e., cases in which the normal coincidence of syntactic and semantic headedness fails.

The notion of headedness, whether syntactic or semantic, admits a formal definition only after its intuitive content is fixed. In the case of syntax, the head of a construction X=Y-Z is the element among Y and Z that licenses the presence of the other. So the head of *black cat* is *cat*, not *black*. Given the intuitive understanding, the syntactic head of a construction may be formalized as: that element of Y and Z whose categorial features project to X (where there must be special clauses e.g. for N-N compounds). The notion of a semantic head is likewise intuitive: whichever of Y and Z takes the other as argument is the head; and if neither does, then (since the only remaining transaction between them is that of θ-identification) it is the element whose θ-position the other identifies with. So the semantic head of *see Mary* is the V *see*, taking *Mary* as argument; and the semantic as well as the syntactic head of *black cat* is *cat*, because we are passing from the genus cat to the species, black cat.

Elementary examples of switch-headedness are seen in *alleged thief*, *known problem*, and the like, where the N is in fact a complement of the Adjective, even though the projection is nominal, not adjectival. Thus an alleged thief is a person x of whom it is alleged: that x is a thief. Likewise with certain adverbials: in *Oedipus eagerly married Jocasta*, the verbal or inflectional projection *t married Jocasta* is an argument of the adverbial form *eagerly* of the semantic head *eager*; and so is the subject *Oedipus*. The complement is an opaque domain, whose truth is nevertheless presupposed, since Oedipus could have eagerly married Jocasta only if he did in fact marry Jocasta; at the same time, it does not follow that he eagerly married his mother.[1]

The phenomenon of switch-headedness shows that there is no uniform map between the syntax and the semantics. At the same time, we expect, and we find, that the map is close to uniform, thus presumably constituting the null hypothesis that the learner brings to the problem of acquisition.

The universality of the principles of semantic combination may be subject to counterexample, or else defensible only by begging the question, which is much the same thing. Thus consider the divergence between languages that do and languages that do not observe the indefiniteness condition in (their versions of) *there*-insertion. The English *there is John is the garden* has only the "list" reading, but the Italian (alleged) equivalent *C'e' Gianni in giardino* does not. Therefore, if indeed the relevant structures are alike except for differences in features inessential to the semantics, universality would fail: the Italian sentence has a combinatorially determined meaning that the English sentence lacks. Of course, one could always make a lexical distinction of some sort between English *there* and Italian *ci*, and put the divergence down to a difference in lexical selection. But this view would have to be accompanied by a reason to believe that the difference ought to make a semantic difference (particularly since, locatives such as *in giardino* apart, Italian also observes the indefiniteness condition). (Moro (1997: 150 ff.) endeavors to supply such a reason, perhaps correctly.) Otherwise, the distinction corresponds to painting one morpheme red and the other blue, and so rescuing universality at the cost of a wholly gratuitous postulation of formal features.

What formal features may we insert? For a simple but problematic case, consider nominal minimal pairs such as *the criminal's capture* vs. *the criminal's observation*. The first can be interpreted as referring to the capture of the criminal, but the second cannot be interpreted as referring to the observation of the criminal; and we know from various work the contours of the generalization governing this distinction (the "affectedness condition," due to Mona Anderson (1979)). To restore compositionality, therefore, we must inject that condition into the syntax; that is, it must be represented amongst the formal features of the head nouns in question (for otherwise there will be no distinction in the permissible interpretations of these phrases, the words *capture* and *observation* both being ordinary nominals derived from transitive verbs). Certainly, one can put the features in: but it should not be supposed that they are "formal" in any serious sense, derived as they plainly are from the semantics of the corresponding verbs. Is that a cheat? The question of universal compositionality, understood as a locality condition on semantics, seems like a game where it is not only possible to cheat, but also where there are no clear rules that will determine in advance whether one is cheating.

3. Conditionals: A Problematic Case.

Of the two alleged examples of non-compositionality that I suggested in Higginbotham (1986), one, namely the suggestion that complement clauses denote only relatively to what they are embedded in, has been suggested also by others in various forms, and much debated. On one version (not the one I gave) a clause such as *snow is white* expresses an intension, considered in isolation, but when appearing with a complementizer, as in *that snow is white*, it expresses (and the complementized clause denotes) an element with internal structure, or "structured meaning," as in Cresswell's work. I will advert to this proposal only in passing, in connection with issues involving anaphora. The other, which I will enlarge upon here, had to do with conditionals, as in minimal pairs like (17)-(18):

(17) Everyone will succeed if he works hard.

(18) No one will succeed if he goofs off.

The observation was that *if*, while interpretable as a conditional connective in (17), could not be so interpreted in (18) (for it would then mean that there is no one whose goofing off would be a sufficient condition for his success). A counter-thesis, which has acquired something of folkloric status, is that in both cases the conditional *if* is meaningless: the clause that it marks constitutes merely the restriction on the quantifier, as made explicit in the paraphrases (19)-(20):

(19) Everyone who works hard will succeed.

(20) No one who goofs off will succeed.

I will argue that this *riposte* fails in general (even if it appears to succeed for these cases), and that the counterexample still stands. But there will be a further moral to the story: compositionality can be restored under certain assumptions about the meaning of conditionals. The compositionality hypothesis will thus have consequences for other matters, most properly involving philosophical logic.

The problem, and the counter-thesis, so far as I know, were first adumbrated in a few swift remarks in Lewis (1975: 14-15). Lewis did not, at least as I read his work, intend the counter-thesis to be universally applicable; that is, applicable to absolutely all conditionals. Rather, it was to apply to cases of what he called "unselective binding," illustrated for instance by Frege's example (21):

(21) If a number is less than 1 and greater than 0, then its square is less than 1 and

greater than 0.

and by many others in Lewis's important article.

In Higginbotham (1986), I observed that the generalization, that *if* and *unless* are interpreted differently depending upon the nature of the higher quantifier, extended to a contrast between all monotone increasing quantifiers such as *every*, on the one hand, and all monotone decreasing quantifiers such as *no* on the other.[2] I was there short on examples, however, which I now supply. Suppose that we are speaking of the 30 students now enrolled in Philosophy 300, and consider (22):

(22) Most students will get A's if they work hard.

where *they* is construed as bound to *most students*. First of all, (22) must be sharply distinguished from (23), the result of absorbing the *if*-clause into the restriction:

(23) Most students who work hard will get A's.

For: (22) is true iff in counting up the students *x* of whom it is true to say that *x* will get an A if *x* works hard, the total amounts to most of them; whereas (23) is true or false depending upon whether, of those students who in fact work hard, most get A's. So (22) and (23) are logically independent. That is enough to show that the absorption method suggested by Lewis for his cases will not work in general; and as we will see below, it fails also even for the universal and negative existential quantifiers. Consider now (24):

(24) Few students will get A's if they work hard.

again with *they* construed as bound to the subject. This example certainly does not appear to mean that few students are things *x* such that *x* will get an A if *x* works hard.[3] We will return to the question what exactly it does signify, but for immediate purposes it is sufficient to note that it does not appear to mean the same as the straightforward (25):

(25) Few students who work hard will get A's.

and, even if it did, the problem for compositionality would remain: for the question how, if at all, the subordinating conjunction *if* is to be interpreted could not be locally determined.

Returning now to the issues posed by the original examples, I want to propose a general account of the conditional that sorts out the phenomena; only afterward will I consider further the question of compositionality.

First, let us reduce the range of data. As remarked above, and noted in my earlier work, the issue of how to interpret the connective arise for *unless* as well as for *if*. Thus we may contrast (26) and (27):

(26) Every student will get an A unless he goofs off.

(27) No student will get an A unless he works hard.

(26) has it that for every student x, x will get an A provided that x doesn't goof off; but (27) cannot be taken as meaning that for no student x, x will get an A provided that x doesn't work hard.

Now, one is taught in Logic 101 to "translate" *unless* by disjunction '$\vee$'; and this is OK, but only as it were by accident. The reason it is OK is that, whereas *p unless q* pretty clearly amounts to *p if not-q*, the schemata '$\neg q$ -> p' and '$p \vee q$' are truth-functionally equivalent. The equivalence of course fails where the *if*-clause does not express the material conditional. Even so, if we take *unless*-clauses as decomposed or else in taken up in some semantic fashion (depending upon how one treats the conditional) as what we might call *if*+*not*-clauses, then the problem posed by (26)-(27) immediately reduces to the previous case, (26) being equivalent to (28), and (27) to (29):

(28) Every student will get an A if he doesn't goof off.

(29) No student will get an A if he doesn't work hard.

Hence I confine the examples in what follows to the conditional *if*.

On reflection, it is clear, I believe, that some but not all universal indicative conditionals of the type that we have been discussing, where the subject universal binds a place in the *if*-clause, can be treated by the method suggested in Lewis. For an example that clearly pulls them apart: suppose that the university is going to offer generous pensions to some 20% of its 422 professors, hoping to induce early retirement, but has not yet decided, or even drawn up criteria for deciding, which 20% this will be. Concluding as I do that generous pensions will infallibly induce early retirement, I believe (30):

(30) Every professor will retire early if offered a generous pension.

(30) of course implies (31), the result of absorbing the *if*-clause into the restriction:

(31) Every professor offered a generous pension will retire early.

But the converse is false: there might be many professors (but even one will do) who we can be sure will not retire early, quite independently of any pension they may be offered.

Similar examples may be constructed for the negative existential, showing that absorption in this case fails too. It may be that I have taken a poll of the professors, determining the truth of (32):

(32) No professor will retire early if not offered a generous pension.

That will imply (33):

(33) No professor not offered a generous pension will retire early.

But again the converse is false: if Professor X is going to retire early, period, then he is a counterexample to (32). But if he is amongst those offered a generous pension, then he is no counterexample to (33), whose truth or falsehood depends only upon whether any of those in the 80% not offered a generous pension retire early.

Examples can be multiplied; but I want now to take more theoretical steps. I will take up the indicative conditional as suggested by Stalnaker (1968): *q if p* is true in *w* iff *q* is true in the closest *p*-world $w' = f(p,w)$ to *w*, or else there are no worlds in which *p* is true. Writing the Stalnaker conditional as '⇒', we have the validity of (CEM), or Conditional Excluded Middle:

(CEM) $(\varphi \Rightarrow \psi) \vee (\varphi \Rightarrow \neg\psi)$

a point that will play a role in what follows.

When is the Stalnaker conditional $\varphi \Rightarrow \psi$ equivalent, for a given world *w*, to the material conditional φ->ψ? I distinguish three cases: (a) $f(\varphi,w)$ is undefined (or, alternatively, is a sink state); (b) φ is itself true in *w* (in which case $f(\varphi,w)=w$, by definition); and (c) $f(\varphi,w)=w' \neq w$ is defined, but the truth value of ψ in *w'* is whatever it is in *w*. Each case will parley into a case where an *if*-clause can be taken as merely restricting a universal quantifier, as follows. Any sentence *Everything is* ψ *if it is* φ, taken as (34):

(34) $(\forall x)\,(\varphi(x) \Rightarrow \psi(x))$

will be equivalent to (35):

(35) $[\forall x\!: \varphi(x)]\ \psi(x)$

whenever, for each object a, either (a') $\varphi(a)$ is false in every world; or (b') $\varphi(a)$ is true in every world; or (c') neither (a') nor (b'), but, where $w' = f(\varphi(a),w)$, the truth values of $\psi(a)$ in w' and w coincide.

Intuitive examples of the equivalence of (34) and (35) include Frege's (21), and any similar mathematical case. Those falling particularly under (c') include examples like (36):

(36) Every book on that shelf is boring if it has a red cover.

One is ready to regard (36) as equivalent to (37):

(37) Every book on the shelf with a red cover is boring.

But that is because we know that giving a book a red cover does not alter its contents, so does not affect whether it is boring. Let b be a book on that shelf with a blue cover. In the closest possible world, whatever it is, in which b has a red cover, it is boring or not, just as it is boring or not as things are.

We are edging here toward the interaction of indicative conditionals with counterfactuals, which would take us far afield. I introduce, however, as a technical device, the notion that φ is *counterfactually irrelevant* to ψ *in w* if either φ is necessarily false, or $\varphi \Rightarrow \psi \Longleftrightarrow \psi$ holds in w; and that φ is counterfactually irrelevant to ψ if irrelevant in w for every w. Extending this notion to open sentences, I will say that $\varphi(x)$ is counterfactually irrelevant to $\psi(x)$ if $\varphi(a)$ is counterfactually irrelevant to $\psi(a)$ for every a. We then have the generalization (I):

(I) If $\varphi(x)$ is counterfactually irrelevant to $\psi(x)$, then (34) and (35) are equivalent, for every w.

(The converse is false, as it may just happen that (34) and (5) are equivalent because $\varphi(a)$ and $\psi(a)$ both fail as things are, whereas both hold in $w' = f(\varphi(a),w)$.) Evidently, mathematical universal conditionals will show counterfactual irrelevance. But so will examples like (36) (assuming it quite impossible that the color of a book's cover should have any influence on whether its contents are boring). On the other hand, it is the counterfactual relevance of $\varphi(x)$ to $\psi(x)$ that points up the difference between (30) and (31), repeated here:

(30) Every professor will retire early if offered a generous pension.

(31) Every professor offered a generous pension will retire early.

To assess the truth value of (30), we must know whether professor Y, who is not offered a generous pension, and does not in fact retire early, would have retired early had she been offered one. Nothing like that is at stake for (31).

Having, if I am right, vindicated the independence of the conditional meaning for cases like (30), we are brought back to the problem posed by the inequivalent (32) and (33). (32) in particular cannot, it would appear, be understood as in (38):

(38) (No x) (x is not offered a generous pension $\Rightarrow$ x will retire early).

But, as not being offered a generous pension is counterfactually relevant to retiring early (for we need to know whether professor X, who did in fact retire early, would have done so had he not been offered a generous pension), (32) is not equivalent to (33) either.

Thus we are brought, I believe, to an obvious hypothesis. We may decompose *for no x, A* as *for all x, not-A*, and note, by the general principle (CEM) that characterizes the Stalnaker conditional, that $\neg(\varphi \Rightarrow \psi)$ is equivalent to $\varphi \Rightarrow \neg\psi$. By this double transformation, (30) is then equivalent to (39):

(39) Every professor will not retire early if not offered a generous pension.

But now in *this* expression, the link between the *if*-clause and the main clause is rightly expressed by '$\Rightarrow$'. But that means that (32) can be understood as in (38) after all!

The trick that we just pulled with *no* can be pulled with any monotone decreasing quantifier. The lexicography of quantifiers, as Frege taught us, is that they map concepts into truth values (or, in natural languages, as Frege also observed, ordered pairs of concepts into truth values; I will confine the discussion here to the unrestricted case, the extension to restricted quantifiers like *every student* being immediate). Recasting this lexicography in model-theoretic terms, a quantifier Q on a non-empty domain D is monotone increasing if for any subsets X and Y of D, $Q(Y)$=True if $Q(X)$=True, and $X \subseteq Y$; and it is monotone decreasing if $Q(Y)$=True if $Q(X)$=True, and $Y \subseteq X$. So for each monotone decreasing Q there is a monotone increasing Q' such that for all X:

$$Q'(D\text{-}X) = Q(X)$$

Consider in this light the puzzling example (24), repeated here:

(24) Few students will get A's if they work hard.

We may decompose *for few x, A* as *for most x, not-A*, and apply (CEM). According to this transformation, (24) should amount to (40):

(40) Most students will not get A's if they work hard.

I am unable to convince myself whether this is the right answer, or (24) is just anomalous: see the "further remarks" below.

I return in any case to the problem of compositionality. What we have seen, if the views advanced here are correct, is that compositionality can be restored, not indeed by making the *if* or *unless* clauses part of the quantifier restriction, but rather by what I have casually called a kind of decomposition and transformation of the sentences in question. For the case of *unless*, this is trivial: if the Stalnaker conditional is assumed, then *q unless p* can be defined by: *q* in the closest possible *not-p*-world. But the decomposition of *no*, for instance, is not trivial. Suppose that the syntactic structure that is the input to semantics for (41) is as in (42) (I use QR, but this is inessential):

(41) No student will get an A if he goofs off.

(42) [[No student]$_i$ [t_i will get an A [if he$_i$ goofs off]]]

Despite initial appearances, the correct compositional result is obtained, but only by exploiting the law (CEM) that characterizes the Stalnaker conditional, a controversial assumption. So perhaps in the end I have offered an argument in favor of the Stalnaker conditional. If so, then that is a kind of consequence of the assumption of local compositionality that I am arguing here should be addressed.

[*Further Remarks*. We may be able to push the analysis farther. Suppose we adopt the interpretation of the indicative conditional suggested in Lewis (1973), and accept the Limit Assumption, so take *p*=>*q* as true in *w* iff *q* holds at every closest world in which *p* is true. Then Conditional Excluded Middle may fail. Define counterfactual irrelevance as above. Then (I) continues to hold. However, we now admit cases in which *Nothing is B if it is A* need not amount to *Everything is not B if it is A*.

The examples that we have given to this point are not examples that, in the most intuitive sense, support Conditional Excluded Middle. Thus we are not ready to say, in typical settings, *Either Professor X will retire early if offered a generous pension, or Professor X will not retire early if offered a generous pension*, or *Either student Z will not succeed if she goofs off, or student Z will succeed if she goofs off*. But there are cases where, again on the most intuitive level, Conditional Excluded Middle applies. Suppose a bowl on the table containing a large quantity of peanuts, enough to supply everyone at the reception. Each particular person is either allergic to peanuts, or else not. So I can volunteer, for each person *x*: either, if *x* eats those peanuts, *x* will have an allergic reaction to them, or, if *x* eats those peanuts, *x* will not have an allergic reaction to them. Now, I know that allergy to peanuts is rare, and so am confident in saying (43):

(43) Few people will have an allergic reaction if they partake of those peanuts.

Obviously, this is to be distinguished from (44):

(44) Few people who partake of those peanuts will have an allergic reaction.

But now the question is whether (43) amounts to (45):

(45) Most people will not have an allergic reaction if they partake of those peanuts.

or, more tendentiously, (46):

(46) There are few people (among those at the reception) such that their partaking of those peanuts is a sufficient condition for their having an allergic reaction to them.

If the answer to this question is affirmative, then we may propose that conditionals *Q things are B if A*, where *Q* is monotone decreasing, are anomalous if Conditional

Excluded Middle fails (for some values of the variable), otherwise equivalent to *Q' things are not-B if A*, where *Q'* is the monotone increasing quantifier corresponding to *Q*. Perhaps conditionals with monotone decreasing quantifiers presuppose (CEM).

What of conditionals with quantifiers that are neither monotone increasing nor monotone decreasing, *exactly three*, for instance, or *between 4 and 17*, or *some odd number of*? It appears to me that the conditional is satisfactorily interpreted as the conditional connective in contexts where (CEM) holds, but is anomalous otherwise. Thus, for instance, it is unclear what *Exactly three students will pass unless they goof off* is supposed to mean, whereas *Exactly three guests will have allergic reactions unless they avoid those peanuts* is pretty clear.

Further investigation would involve overt counterfactuals, which I do not consider here.]

4. Linguistic Relations.

I consider now how to extend the notion of local compositionality so as to take account of the fact that the features of a syntactic structure include not only the formal features of points in that structure (these are, in effect, predicates true of the points: for the earliest exposition, see Chomsky (1975: 177) on the "*X* is a *Y*" relation), but also relations between the points. I will confine myself to a particular case, where the relations in question are all binary, and may be thought of as relations of *immediate antecedence*; i.e., a relation of antecedence *A*->*B* where there is no *C* such that *A*->*C*->*B*, and where *A* is to be thought of as inheriting its (entire) interpretation from *B*. The elements having antecedents are, I will assume, to be taken as free variables; i.e., in structures in which their antecedents have so to speak not yet appeared, they will have values on *assignments* of some appropriate sort. If we are dealing with humdrum cases of pronominal anaphora, as in *John loves his mother*, with *his* anaphoric to *John*, then the value of the pronoun *his* on an assignment can be any thing. But in VP-deletion, as for instance in *John's mother doesn't* ∅, the interpretation is ultimately recovered through an expression: as Tanya Reinhart pointed out long ago, we must be able to distinguish *John wants to become a doctor, but his mother doesn't want that* (i.e., doesn't want him, or perhaps herself, to become a doctor) from *John wants to become a doctor, but his mother doesn't* ∅, whose second clause means only that his mother doesn't want to become a doctor. To put it another way, VP-deletion appears to be a "copying" rule, and it is perhaps simplest if feasible to suppose that copying is done before interpretation comes along.

One way of thinking of the syntactic trigger for anaphoric interpretation is through assignment of *indices* to anaphoric and other elements, as originally suggested in Chomsky (1965); or, and perhaps more perspicuously, the assignment of some set of equivalence-relations to the points in the syntactic tree. On this procedure, an assignment will be an assignment of values to a given index; the semantics will then be constrained through a system of equations, and we may call an assignment *appropriate* for an index just in case the value of all elements with that index is the same. So in $John_i$ *loves* his_i *mother*, for example, the reference of $John_i$ on any assignment is John; that of his_i is whatever the assignment a assigned to i; and the condition to be satisfied for appropriateness is that $a(i)$=John.

The indicial view, however, does not distinguish between those cases where *A* and *B* are intended to have the same reference (and intended to be understood as so intended) from those cases in which it is intended that *A is to have* the reference of *B*. As Howard Lasnik originally observed, the distinction between these cases correlates with the availability of "sloppy identity;" i.e., with the distinction between *His father hates John, but not Bill*, or *His father hates John, but the man doesn't hate Bill*, where the reference of *his father* must be preserved (so the identity is rigid), and *John hates his father, but Bill doesn't*, or *John hates his father, but Bill doesn't hate him*, where the reference may shift (so the identity is flaccid). (Or try: *USC loves its football team, while MIT is indifferent to it*.)

On the indicial view, it is easy to stipulate that an assignment be appropriate; and the stipulations can be carried out piecemeal, moving interpretively up the tree, as originally outlined in work by Jon Barwise from some years back (Barwise (1987)). But if, as I suggest, anaphoric relations are asymmetric, then we shall want to track, for instance in the case where *Z* contains an element whose immediate antecedent lies in *Y*, the effect on the meaning of *Z* as it enters the construction *X*=*Y*-*Z*. We can certainly make it a formal feature of a point *P* in **T** that it has an immediate antecedent somewhere in **T**: this much "global lookup" has to be tolerated so as to distinguish anaphoric from non-anaphoric uses of the same forms. But as this antecedent may be arbitrarily far away from *P*, it may not be visible, not only to *P* itself, but also to any number of constituents properly containing *P*. Hence, appropriateness of assignments must be tracked independently, as meaning is built up. For the case we are considering, it appears sufficient to specify (47):

(47) a is appropriate for $\mathbf{T}=\mathbf{T_1}\text{-}\mathbf{T_2}$ iff (i) a is appropriate for $\mathbf{T_1}$ and $\mathbf{T_2}$, and (ii) for every α in **T** having an immediate antecedent *W* in **T**, $a(\alpha)=M(W)$.

The rule (47) does exclude some syntactic conditions that generally prohibit anaphoric relations; e.g., the condition that an anaphoric element cannot seek its antecedent within its own c-command domain. Nor does it rule out cases of "circular reference," such as have been discussed recently by Polly Jacobsen. Thus "circular" examples such as *His wife saw her husband* receive truth values on appropriate assignments, specifically those a that assign to *his* whatever the reference of *her husband* is on a, at the same time assigning to *her* on a whatever the reference of *his wife* is on a. However, the reference to assignments being ineliminable, no truth value can be determined.

It is of course easy to construct hypothetical languages for which the appropriateness of an assignment of values to variables was not governed by the elementary rule (47). So I would like to suggest that (47) is an appropriate counterpart to compositionality, expressing the local compositionality of anaphora.

There is a semantical issue here, discussed in Dummett (1973) with reference to Frege. Consider (48):

(48) John thinks Mary is a spy, but she isn't.[4]

where *she* picks up its interpretation from *Mary*. On Frege's view, the term *Mary* has indirect reference in (48); hence the anaphor *she*, if simply picking up its reference from its antecedent, should refer to that as well. But of course it refers just to Mary. On a view such as that in Higginbotham (1986), we would resolve the issue this way: the expression *Mary* has one value considered within the complement clause *Mary is a spy*, namely Mary herself, and another when considered as an element of the whole sentence *John thinks Mary is a spy*, namely the word *Mary*, understood as if the speaker had said it. To get the right result, we have to look at the value of *Mary* within the *least* element containing it (where compositionality holds); that is, its own clause. In a certain sense, the anaphora are then non-compositional, because it is not the meaning pure and simple, but the relativized meaning that must be consulted. But of course my account was going non-compositional anyway, and the non-compositionality of the anaphora follows from this.

To sum up: I have urged that the compositionality of semantics, either in the way I have defined and illustrated it, or in some other way, be seen neither as a triviality, nor as a quasi-conceptual issue, but rather as an empirical hypothesis, right or wrong, about the projection of interpretation in human languages. A good way to look at the subject, I think, is neither to insist upon compositionality at all cost, nor to propose that exactly this or that syntactic information is available to the interpretive component of a grammar; but rather to take compositionality in a strong form as a working hypothesis, and see what the consequences are. In this way, I hope, the question of the scope and limits of compositionality can be deflected from what seem to me the excessively *a priori* pronouncements that might otherwise be made about it.

Notes

*Various of the themes in this paper figured in a presentation at the Montreal meeting on Asymmetry in Syntax, 2001, and in talks at the University of Arizona and at UCLA, the latter as part of the Joint Syntax/Semantics Workshop between USC and UCLA, also in 2001. I am grateful to the organizers of these occasions for the opportunities that they afforded, and especially to Anna-Maria Di Sciullo and her colleagues for their hospitality in Montreal. This particular draft is for the Michigan meeting on Semantics, to be held at Ann Arbor, Michigan, November 2002, Paul Pietroski and Ernest Lepore to comment. I shall be pleased to receive any communication concerning its contents, either before or after the meeting.

1. For details on this and similar examples, see Higginbotham (2000: 57-60).

2. A restricted quantifier is monotone increasing if *Q A are B* and *All B are C* implies *Q A are C*, and monotone decreasing if *Q A are B* and *All C are B* implies *Q A are C*. See further below for the model-theoretic version of these notions.

3. Although I will not expand upon the point here, it should be noted that

(i) *Q* things are *A* if they are *B*

and

(ii) *Q* things are things such that they are *A* if they are *B* need not be equivalent in this setting. In examples of the form (ii), the conditional always carries its proper force; i.e., the force it would have in *x is A if x is B*; so there is no problem about compositionality. To put the matter in Lewisian terms, the quantifier in (ii) is a whole clause away from the *if*-clause, so absorption of it into the quantifier restriction is blocked, presumably for syntactic reasons. Compositionality is an issue only for examples of the form (i).

4. I am indebted to Gabriel Segal for the example, and discussion of the issues here.

References

Anderson, Mona (1979). *Noun Phrase Structure*. Unpublished doctoral dissertation, University of Connecticut, Storrs, Connecticut.

Barwise, Jon. (1987). "Noun Phrases, Generalized Quantifiers, and Anaphora."

Gärdenfors, Per (ed.), *Generalized Quantifiers*. Dordrecht, Holland: D. Reidel. pp. 1-30.

Chomsky, Noam (1975). *The Logical Structure of Linguistic Theory*. Chicago: The University of Chicago Press.

Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The MIT Press.

Cresswell, Max (2002). "Why Propositions Have No Structure." ms.

Dummett, Michael (1973). *Frege: Philosophy of Language*. London: Duckworth.

Gazdar, Gerald; Klein, Ewan; Pullum, Geoffrey; and Sag, Ivan (1979). *Generalized Phrase Structure Grammar*. Cambridge, Massachusetts: Harvard University Press.

Heim, Irene and Kratzer, Angelika (1998). *Semantics in Generative Grammar*. Malden, Massachusetts: Blackwell.

Hendriks, Herman (2001). "Compositionality and Model-Theoretic Interpretation." *Journal of Logic, Language and Information* **10**, 1. pp. 29-48.

Higginbotham, James (1985). "On Semantics." *Linguistic Inquiry* **16**, 4. pp. 547-593.

Higginbotham, James (1986). "Linguistic Theory and Davidson's Program in Semantics." LePore, Ernest (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell. pp. 29-48.

Higginbotham, James (2000). "On Events in Linguistic Semantics." Higginbotham, James; Pianesi, Fabio, and Varzi; Achille (eds.), *Speaking of Events*. New York: Oxford University Press. pp. 49-79.

Hodges, Wilfrid (2001). "Formal Features of Compositionality." *Journal of Logic, Language and Information* **10**, 1. pp. 7-28.

Krifka, Manfred (1992). "Thematic Relations as Links Between Nominal Reference and Temporal Constitution." Sag, Ivan A. and Szabolcsi, Anna (eds.), *Lexical Matters*. Stanford: CSLI Publications. pp. 29-53.

Lewis, David (1973). *Counterfactuals*. Cambridge, Massachusetts: Harvard University Press.

Lewis, David (1975). "Adverbs of Quantification." Keenan, Edward L. (ed.), *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press. Reprinted in Lewis, *Papers in Philosophical Logic*, Cambridge University Press, Cambridge, 1998, pp. 5-20."

Moro, Andrea (1997). *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge: Cambridge University Press.

Stalnaker, Robert (1968). "A Theory of Conditionals." Rescher, Nicholas (ed.), *Studies in Logical Theory*. Oxford: Basil Blackwell. pp. 98-112.