

# THE SHORT ANSWER: IMPLICATIONS FOR DIRECT COMPOSITIONALITY (AND VICE VERSA)

PAULINE JACOBSON

*Brown University*

This article is concerned with the analysis of ‘short’ or ‘fragment’ answers to questions, and the relationship between these and the hypothesis of DIRECT COMPOSITIONALITY (DC) (e.g. Montague 1970). DC claims that the syntax and semantics work ‘in tandem’ to prove expressions well formed, while at the same time assigning them a meaning (a model-theoretic object). DC makes it difficult to state any kind of identity condition for ‘ellipsis’ and would hence lead one to suspect that short answers do not contain hidden linguistic material. This article argues that they indeed do not. Rather, as proposed in Groenendijk & Stokhof 1984, the question and short answer together form a linguistic unit, which I call a Qu-Ans, whose semantics gives the proposition that is understood as following from the pair. Three new arguments are adduced for the Qu-Ans analysis over one making use of silent linguistic material, and a core class of traditional arguments for silent linguistic material are answered. Moreover, it is shown that many of the traditional arguments for silent linguistic material themselves presuppose a non-DC architecture. If (as is claimed) these arguments do not hold, the Qu-Ans analysis of short answers actually supports the DC view, under which no use is made of logical form, and no use is made of representational constraints on structure.\*

*Keywords:* direct compositionality, fragment answers, ellipsis, case matching, binding

**1. INTRODUCTION.** This article addresses both a big question and a more local one. The big question centers on the viability of the HYPOTHESIS OF DIRECT COMPOSITIONALITY, put forth in, for example, Montague 1970 and many works since. This is that the syntax and semantics work in tandem, rather than the syntax being an autonomous system that computes well-formed representations that serve as inputs to semantic interpretation. Put another way, direct compositionality assumes that the syntax proves the well-formedness of expressions in a language (often proving larger expressions well formed on the basis of smaller ones), while the semantics works in tandem—assigning each local expression a meaning as it is ‘built’ (i.e. proven well formed) in the syntax. Moreover, we will take the meaning (or ‘semantic value’) of an expression to be a model-theoretic object and not a symbolic representation, although I use symbolic representations as a way to name those objects.<sup>1</sup> Direct compositionality is an extremely simple view of the architecture of the grammar. It makes no use of distinct levels of representation and in fact (under a particular conception to be elaborated below) actually makes no use of ‘representations’ in the statement of rules. And since any theory needs a compositional syntax and any theory needs a compositional semantics, having the two work together would seem to be the null hypothesis. All other things being equal, then, it should be a welcome result whenever a construction that has been taken to require a

\* For helpful comments and discussion, I would like to thank Scott AnderBois, Greg Carlson, Patrick Elliot, Donka Farkas, Laura Kertz, Jason Merchant, Geoff Pullum, Jeff Runner, Ivan Sag, Barry Schein, Balasz Suranyi, Anna Szabolcsi, and three anonymous referees, as well as audiences at the 2009 Rutgers Semantics Workshop, the 2009 HPSG Conference (Göttingen), the Stanford Ellipsis’Event (2011), the 9th Tbilisi Symposium (2009), and at colloquia at NYU, Yale, UMass, Rochester, and Edinburgh. This work is supported by NSF Grant BCS0646081.

<sup>1</sup> Note then that this distinguishes the direct compositional program advocated here from some current versions of MINIMALISM that look closer (in spirit at least) to direct compositionality in that in some versions the syntactic representations and the logical forms (LFs) are computed ‘in tandem’. But if LF is not the final semantic object but is instead the input to the model-theoretic interpretation (as is usually assumed), this in fact is still not a direct compositional theory in the sense here.

non-direct compositional analysis can be shown to be compatible (without undue complexity) with direct compositionality. In the case at hand (fragment answers), I show that—contrary to much received wisdom—a direct compositional analysis not only is possible, but is also actually preferable to at least one rather standard non-direct compositional alternative.

Thus—as anticipated in the above remarks—the more local question revolves around a case study: the analysis of ‘short’ (or fragment) answers to questions, as in the dialogue in 1.

- (1) a. Q: Who left the party at midnight?  
 b. A: Claribel.

In particular, does 1b contain silent or deleted material such that it is really (at some level) the sentence *Claribel left the party at midnight*? Let’s call this the SILENT LINGUISTIC MATERIAL hypothesis, or SLM for short. The answer that I argue for (in short) is ‘No’. The claim that there is no silent/deleted material has also been made by many previous researchers; see, for example, Culicover & Jackendoff 2005 for this claim with respect to a more general set of fragments.<sup>2</sup> I confine the current study to fragment answers only. The analysis advocated here is based heavily on one put forth originally by Groenendijk and Stokhof (1984) and—in a somewhat different form—by Ginzburg and Sag (2000). While the current analysis differs in details from either of these, it is in the same spirit.

The article is structured as follows. I first elaborate on the hypothesis of direct compositionality and show why the analysis of fragment answers bears on its feasibility (§2). The analysis of fragment answers to be defended here is developed in §3, and §4 presents three arguments for the analysis here over the SLM view. I then answer a sample of arguments that have been presented in favor of SLM. Space of course precludes addressing every single argument that has ever been advanced for SLM, but it is hoped that the representative sample given in §5 will serve to cast doubt on the claim that there is strong evidence for SLM. In particular, many arguments for SLM are based on rather entrenched analyses of various phenomena (e.g. pronominal binding). But these analyses themselves are not viable if direct compositionality is correct, and indeed there are alternative (independently motivated) direct compositional analyses of the phenomena in question. (See also the online appendices for discussion of two additional arguments for SLM.<sup>3</sup>) Section 6 then turns the usual arguments for SLM on their head: since many are predicated on non-direct compositional analyses of various phenomena, the arguments AGAINST SLM adduced in §4 in turn show that the non-direct compositional analyses of the relevant phenomena cannot be correct. Concluding remarks and further observations are given in §7.

**2. THE HYPOTHESIS OF DIRECT COMPOSITIONALITY.** Let us clarify the larger theoretical question at stake, and why the analysis of fragment answers is relevant. The hypothesis of DIRECT COMPOSITIONALITY (hereafter DC) was put forth in Montague 1970 and has been explored and/or maintained by many since. For example, it was assumed or at least taken as a desideratum in much of the research in semantics in the 1970s and 1980s under the rubric of MONTAGUE GRAMMAR; it is maintained in GENERALIZED PHRASE STRUCTURE GRAMMAR (Gazdar et al. 1985) and in HEAD-DRIVEN PHRASE

<sup>2</sup> See also the extensive body of work by Stainton (e.g. Stainton 2005 and works cited there) for a discussion of non-SLM accounts of fragments more generally.

<sup>3</sup> The online appendices referenced throughout this article can be accessed at <http://muse.jhu.edu/article/619541/pdf>.

STRUCTURE GRAMMAR (see e.g. Pollard & Sag 1994), and in much of the work in CATEGORIAL and TYPE LOGICAL GRAMMAR. Under this view, one can think of every linguistic expression (a word, a sentence, or any phrase in between) as a triple consisting of ⟨sound, syntactic category, meaning⟩, where by ‘meaning’ I intend a model-theoretic object and not a symbolic representation. I use the notation  $[\alpha]$  to indicate the sound of an expression  $\alpha$  (although, for convenience, I use standard orthography and not phonetic representation) and  $\llbracket \alpha \rrbracket$  to indicate the meaning of  $\alpha$ . The rules of the grammar take one or more of these triples as input, and each rule yields a triple as output.

For example, a familiar phrase structure rule such as  $S \rightarrow NP VP$  is an abbreviation for the phonological and syntactic parts of the combinatory rule shown in 2; the third part would specify the meaning of the output expression in terms of the meanings of the two inputs (i.e. the two daughters NP and VP). Assuming that VPs denote functions from individuals to truth values and NPs denote individuals, the particular semantics is as shown in 2.

- (2) Given an expression  $\alpha$  of the form  $\langle [\alpha], NP, \llbracket \alpha \rrbracket \rangle$  and an expression  $\beta$  of the form  $\langle [\beta], VP, \llbracket \beta \rrbracket \rangle$ , there is an expression  $\gamma$  of the form  $\langle [\alpha-\beta], S, \llbracket \beta \rrbracket(\llbracket \alpha \rrbracket) \rangle$ .

Of course, nothing in this general setup commits us to the view that the grammar contains many particular rules like that in 2; this rule is given just for illustration. Rather, the rules may be listed instead as very general rule schemata. (See, for example, CATEGORIAL GRAMMAR for one way to do this.)

The important point is that the rules prove certain strings to be well-formed expressions of a given category and simultaneously assign them a meaning. Often—as in the case where two expressions combine—larger expressions are proven well formed on the basis of smaller ones, hence the metaphor of the syntax ‘building’ expressions. Of course, this very general picture needs to be supplemented with some hypothesis as to just what sorts of operations are available in the syntax. I take a rather impoverished view of what the syntax can do: assume that when two or more expressions combine they do so only by concatenation (as in 2) or by infixation of one expression into another (the latter being what has been dubbed ‘Wrap’ in the categorial grammar literature; see e.g. Bach 1979, Dowty 1982, and many since). Nothing in this general picture precludes the existence of a unary rule—that is, a rule that takes a single triple as input and yields a triple as output; we return to this point below.<sup>4</sup>

This view has several interesting consequences. First, each local expression that is well formed according to the syntax has a meaning—there is no need to consult surrounding material in order to assign a meaning. Second, there is no use of an abstract level such as LOGICAL FORM (LF); the semantics is computed as the syntax proves well-formed surface (pronounced) expressions. In fact, while DC is sometimes discussed under the rubric of ‘surface interpretation’, this is misleading—the input to the semantics is not any level of representation. Which brings us to the third and most central point to the discussion here: the grammar under this view does not actually make any reference to structural properties at all. In sloganistic terms, the grammar does not ‘see’ structure.<sup>5</sup> A structure such as a familiar tree can, for example, provide a convenient

<sup>4</sup> Note that a unary rule can always be ‘traded in’ for an empty operator, and vice versa. In fact, it is not clear that there is any empirical difference between a unary rule and an empty operator, or whether these are just different metaphors for the same thing.

<sup>5</sup> A caveat is in order here. If one adopts the view that there are ‘Wrap’ (infixation) operations in the syntax, then a small amount of structure is necessary for the grammar to keep track of, for it must keep track of some kind of infixation point at which the infixed material is inserted. But this is the only amount of structure needed; a full representation such as a tree is not something the grammar needs any access to.

way of showing how constituents combine to form a larger constituent, but a tree structure itself under this view has no theoretical standing. (A tree is also a rough representation of how the compositional semantics puts meanings together. It is ‘rough’ in that it might, for example, show that two meanings combine to give a third, but it does not specify exactly how the two combine.)

As a consequence, no phenomenon could be accounted for by a principle in the grammar that is stated as a constraint on representations. Consider, for example, the following kinds of statements, often used to account for ‘principle A’ effects, ‘principle B’ effects, and the distribution of bound pronouns.

- (i) An anaphor must be locally c-commanded by a coindexed NP.
- (ii) A pronoun may not be locally c-commanded by a coindexed NP.
- (iii) A binder must c-command a pronoun that it binds (at LF).<sup>6</sup>

(Note that throughout this article I use the term ‘NP’ rather than ‘DP’; the reader who prefers the latter may make the obvious substitutions as needed.) None of these statements—nor any straightforward reconstruction of them—are possible under the DC architecture sketched above. The phenomena for which these principles are designed still must be accounted for, of course, but I argue later that there are alternative accounts of the same phenomena that are at least as simple as adopting principles in the grammar such as (i)–(iii). This in turn has consequences for many of the standard arguments for the SLM approach to fragment answers, for many of those arguments simply assume the existence of these kinds of constraints on representation.

The hypothesis of DC bears on the analysis of fragment answers in another important way. More generally, consider the implications of DC for the analysis of any construction thought to involve deletion/phonological silencing of material UNDER IDENTITY with something else. Although this article is not primarily concerned with VP ELLIPSIS (VPE), it is considerably easier at this point to use VPE for exposition. (The term ‘ellipsis’ is used in a theory-neutral way.) Consider 3.

- (3) Bode can ski that course in three minutes. Lindsay can, too.

There are many different SLM analyses of VPE on the market, but the key claim of all of them is that at some level the second sentence (*Lindsay can, too*) contains a deleted or silent instance of the VP *ski that course in three minutes*. Following general convention, I represent the notion of silent or deleted material using strikethroughs; thus the second sentence under SLM is *Lindsay can ~~ski that course in three minutes~~ too*. The conditions for the silencing or deletion vary from analysis to analysis, but these are always taken to involve some sort of IDENTITY with other linguistically overt material in the discourse context. The requisite condition may be formal identity, roughly as in Sag 1976, which posits that VPE requires formal identity at LF.<sup>7</sup> Other accounts assume that the right notion is semantic identity: the silencing/deletion of the VP is allowed by virtue of there being an overt linguistic VP with the same meaning. This is roughly the account of ellipsis in Merchant 2001. While we return to this in §4, for now the discussion is phrased in terms that are neutral between a syntactic and a semantic identity condition.

<sup>6</sup> The last of these (‘A binder must c-command a pronoun that it binds (at LF)’) actually has a different status from the other two, for given the usual assumptions about how pronominal binding works, it need not be stated as any extra principle in the grammar. Rather, it can be taken simply as an empirical observation: binding would not otherwise be possible under the standard system. But the view of binding from which this follows is itself incompatible with DC since it requires reference to a level of LF. A different and DC view is explored in §5.1; such a view requires no reference in the grammar to ‘binders’ and ‘bindees’.

<sup>7</sup> I say ‘roughly’ here since Sag’s actual account required formal identity up to alphabetic variance, where he gave a rather complex definition of what it means for two formulas to be alphabetic variants.

Consider how VPE might be formulated in the DC framework. One might formulate a rule in the terms here approximately as follows (the input VP presumably has some feature to ensure that it is only VPs selected by auxiliaries that input this rule).<sup>8</sup>

- (4) Given an expression  $\alpha$  of the form  $\langle [\alpha], VP, \llbracket \alpha \rrbracket \rangle$ , there is an expression  $\beta$  of the form  $\langle [\emptyset], VP, \llbracket \alpha \rrbracket \rangle$ , provided that there is some expression in the discourse context whose meaning is identical to that of  $\alpha$ . (Note that one might substitute ‘form’ for ‘meaning’ here.)

The first part of this rule is unobjectionable. Phonological suppression per se is not the problem. The formulation in 4 expresses as syntactic a unary rule in which the difference between the input and the output is only that the phonology of the VP is empty in the output, and this part is entirely compatible with DC. It is the ‘provided that’ part that is the difficulty. Being identical to some other overt material in the discourse context is not a local property of any expression. It is not even a property that can be stated at the level of a sentence, so it is not compatible with DC.

To be sure, it might be possible to formulate some sort of identity condition within a DC theory once we observe that there are other expressions whose values depend on the discourse context. The prototypical examples are indexicals like *I* and *you*, and so one might try to extend the techniques that have been used for indexicals for the case here. To illustrate one way to handle indexicals in general, in the theory put forth in Kaplan 1989, the semantic value of any expression has both what Kaplan calls a CHARACTER, which is a function from speech contexts to the familiar model-theoretic objects (propositions, individuals, etc.), and a CONTENT, which is the value of this function at the relevant speech context. (Hence, the character of a sentence such as *I love to ski* is a function from speech contexts to propositions.) The proposition that this function delivers is the content, and it depends on who is the speaker in the context of utterance. (We return to this treatment of indexicals in §4.4.) Hence, since there are in any case expressions whose value depends on the discourse context (and there are tools for encoding this as part of its ‘local’ meaning), perhaps 4 can be refined in such a way as to allow silencing only in a discourse context in which there is some other linguistically overt VP with the same ‘content’. I leave it to the interested reader to provide such a formulation. Nonetheless, even if this is possible, it seems implausible. First, for the case of indexicals, it is their VALUE that is sensitive to speech context—but here it is simply the EXISTENCE of a ‘silent’ VP that needs to have this sensitivity. Second, the properties of discourse context required to set the value of various expressions do not include facts about actual linguistic utterances. Rather, the value of expressions might depend on speech time, who is the speaker, who is the hearer, and what entities are salient (this is needed for anaphora). But expressions whose value is set by context usually do not care about what has literally been said.<sup>9</sup>

<sup>8</sup> As noted in n. 4, one could use instead an empty operator that combines with the VP to suppress its phonology. Indeed, this is exactly the tack taken by Merchant; his proposal with respect to short answers is elaborated on below. Note, then, that the difference between the use of a unary rule and the use of an empty operator is not significant; the important difference between a DC approach and SLM concerns the viability of an identity condition. This is discussed below.

<sup>9</sup> So, for example, compare the putative identity condition in 4 to the conditions needed for deaccenting of ‘old’ or ‘given’ information. As is well known, material can be deaccented if it is ‘given’ in some sense in the discourse context, but this does not mean it has to be overtly named. It can be inferred in other ways, as in the famous example from Lakoff 1971.

(i) John called Mary a Republican and then SHE insulted HIM. (*insulted* deaccented)

Of course, saying that VPE does not require any identity condition is all well and good, but a proponent of this view must also provide some analysis of just how it is that the discourse in 3 is generally understood in the way that it is. Since VPE is not our primary concern I do not discuss it any further here, but see Jacobson 2003, 2007 for one analysis within a fully DC framework. We now return our gaze to fragment answers; the undesirability of an identity condition under DC follows for the same reason. As is argued below, the requisite identity condition under SLM is actually extremely difficult to state, whereas a fully DC analysis requiring no such condition is quite simple.

### 3. FRAGMENT ANSWERS.

**3.1. THE SLM ANALYSIS.** The SLM analysis of fragment answers is initially best illustrated with the case of a WH-question involving an object, as in the dialogue in 5.

- (5) a. Q: Who did Bozo invite to the party?  
 b. A: Claribel.

Roughly, the idea here is that 5b is—at some level or representation—*Bozo invited Claribel to the party* or *Claribel, Bozo invited t to the party*, where the material *Bozo invited (t) to the party* is deleted or silenced under identity with a portion of the material in the question. Since it will be helpful to have a more explicit analysis to refer to, I briefly sketch the one given in Merchant 2004. According to this analysis, the derivation of the answer involves a two-step process: *Claribel* is fronted to some position that Merchant calls the Spec position of a node labeled FP, and the silenced material is thus a constituent that (ignoring questions about the verbal morphology) is identical to the material following *did* in the question (5a). Moreover, the head of FP is a silent operator, dubbed E by Merchant, which I relabel Shh, and which operates on its sister TP to suppress the phonology of the sister. My relabeling is not simply to be cute. Merchant uses E in other constructions, but we will see in §4.2 that the silencing operator in an FP—or an FP itself—has a dedicated rather than a more general distribution. Both FP and Shh will thus be given their own names here, and later it will be shown that Shh is not just part of a more general silencing operation. (See especially the discussion on focus in §4.2 and the concluding remarks in that section.)

Thus in greater detail, the SLM structure for 5b is as in 6.

- (6) [<sub>FP</sub> [<sub>SPEC</sub> Claribel]<sub>2</sub>] [<sub>F'</sub> Shh [<sub>TP</sub> ~~Bozo invited t<sub>i</sub> to the party~~]]

Shh not only suppresses the phonology of but also imposes a requirement on its sister. Departing a bit from Merchant's precise implementation, let me here just say that Shh requires its sister to be such that there is overt linguistic material somewhere in the discourse context that is identical in some sense (left open until §4.4) to the complement of Shh.<sup>10</sup>

It is worth noting that—modulo exactly the problem discussed above regarding the nonlocality of the identity condition—this could all be recast in a DC framework by

<sup>10</sup> While not made explicit in Merchant 2004, Merchant's actual account requires his E (here 'Shh') to be a kind of indexical: it is an identity function on certain propositions, but the domain of this function depends on the discourse context. Moreover, it must make crucial reference to LINGUISTIC properties of the discourse context. Thus Shh is the identity function on propositions p that are 'e-given', where the definition of e-given is as follows: an expression A is e-given iff 'there is an antecedent A that entails E and that is entailed by E, modulo  $\exists$ -type-shifting'. Note that 'there is an antecedent' has to mean there is an antecedent in the relevant discourse context, and the notion of an 'antecedent' itself (not defined here) has to mean some proposition p' that is the meaning of overt linguistic material. It cannot mean that p' is simply part of the common ground; the entire program of SLM requires the 'antecedent' to have linguistic properties, and so the notion of an 'antecedent' needs to be defined as a meaning that is the meaning of some overt expression. Hence the definition of the domain of Shh requires reference to the linguistic properties of the surrounding discourse and is not just a purely semantic property of the possible inputs to the identity function.

viewing Shh as a unary operator. To be consistent with the general DC assumptions sketched above, I make use of a categorial grammar framework that makes no use of movement or traces. To avoid irrelevant complications, I simplify the example to *Claribel, Bozo invited*. Using the theory of ‘extraction’ developed in, for example, Steedman 1987, the subject *Bozo* and the transitive verb *invited* can directly combine (no object need be introduced) to give the expression *Bozo invited*, whose category is  $S/RNP$ . This notation indicates a category that could combine with an NP to its right to give a sentence. Moreover, in this account, its meaning is  $\lambda x[\text{Bozo invited } x]$ . One could then add a rule (call it Shh) that maps this  $S/RNP$  to an expression with the same meaning and category, but with no phonology. This can then combine with the overt (‘fragment’) NP *Claribel*, where the semantics applies the function above to the individual Claribel, giving the proposition that Bozo invited Claribel. (Although not directly relevant for fragment answers, the reader might wonder about the account of overt topicalization as in *Claribel, Bozo invited*. Here one might assume that there is a unary rule allowing the (overt) expression *Bozo invited* of category  $S/RNP$  to shift to a (homophonous and synonymous) expression of category  $S_1NP$ —that is, *Claribel* would then occur to the left of the overt material *Bozo invited*.) But once again, the devil is in the statement of the identity condition: the input to Shh on this analysis must be restricted to expressions that are identical in some way to other linguistic material in the context, and this of course is not a local property of the expression in question.

Finally, in Merchant’s implementation, the case of a subject question/answer pair (as in 1) also involves fronting the subject to Spec of FP. The DC reconstruction of that would simply allow any  $S/NP$  (regardless of whether it wants an NP to its left or to its right) to input Shh. Thus the basic idea of a Shh ‘empty operator’ or unary rule is unproblematic for a DC framework—it is, as noted above, the statement of the identity condition that is difficult.

**3.2. THE QU-ANS ANALYSIS: WH-QUESTIONS AND ANSWERS.** Surprisingly, no such identity condition is needed, for there is an alternative that I now spell out. The particular analysis to be argued for is a variant of one put forth originally by Groenendijk and Stokhof (1984). Ginzburg and Sag (2000) propose a very similar analysis, though within a somewhat different framework for the semantics. In terms of the syntax, the answer in both 1b and 5b is just the NP *Claribel*, and its semantic value is the individual Claribel.

But we cannot of course leave the story here: we need to predict that a listener hearing the discourse in 1 understands the answerer to be conveying the proposition that Claribel left the party at midnight, whereas in 5 the answerer is conveying the proposition that Bozo invited Claribel to the party. To accomplish this, following the line of analysis first put forth by Groenendijk and Stokhof (1984), I assume that there is an actual linguistic construction I call a QUESTION-ANSWER PAIR (hereafter referred to as the QU-ANS ANALYSIS). Like any other linguistic expression, a Qu-Ans has both a syntax and a semantics. The idea of a discourse-level construction as an actual full-blown grammatical object (with a syntax and semantics) is not commonly found in formal approaches, but there is no special reason to think that the largest unit the grammar has anything to say about is a sentence. Indeed, if there are full-blown linguistic objects beyond single sentences that are governed by grammatical principles, the notion of a question and answer pair surely seems like a very good candidate for such a thing.<sup>11</sup>

<sup>11</sup> There is one interesting way in which a Qu-Ans differs from other expressions like NPs, Ss, and so forth: it is an expression shared across two speakers. As is shown below, the meaning of the Qu-Ans is derived by

Hence we assume that question-answer pairs themselves are such full-blown linguistic objects with a syntax and a semantics (to be discussed momentarily). See especially Ginzburg & Sag 2000 for a detailed discussion and defense of the notion of a Qu-Ans construction.

An immediate objection one might have is that this analysis requires a new construct in grammatical theory (the notion of a Qu-Ans), and so one might also immediately conclude that SLM has an advantage here in that it does not need this notion. But, in fact, this conclusion is too hasty. While SLM might not literally need this particular notion of a ‘Qu-Ans’, I show later that it nonetheless does need some notion of an ‘answer’ to a given question as part of the grammatical machinery. So, it has no advantage over the proposal defended here. And in fact it will be argued that under SLM there actually is no obvious way to define the requisite notion of ‘answer’. I postpone further discussion of this important point until after the current analysis has been developed.

Quite crucial to the analysis is that a Qu-Ans has a syntax as well as a semantics: it is only if a particular syntactic condition is met that this counts as a true Qu-Ans in the sense of a grammatically defined construction. The syntactic requirement is simple: a pair consisting of a question (call that category Qu) and a fragment (which can be of any appropriate category) constitutes a Qu-Ans only if the category of the second member matches the category of the WH-expression in the question. This might seem like a clear violation of the kind of locality restrictions discussed above because the WH-phrase (and hence the information about its category) will be internal—as the leftmost daughter—to the root Qu node. So how can the fragment and the Qu know they ‘fit together’ appropriately? Are we not cheating by having the grammar refer to the internal structure of the WH-question in order to determine the category of the WH-word? The answer is no, for there are many well-known techniques for ‘passing up’ the information of the category of the WH-word to the root node of the Qu. (See, for example, the literature within generalized phrase structure grammar.) Thus in both 1a and 5a the root node of the question can be Qu[NP]. In other words, the grammar composes a Qu-Ans by combining an expression of category Qu[X] with an expression of category X. Incidentally, in categorial grammar it is tempting to recast the category label ‘Qu[X]’ as Qu-Ans/X, by which the syntax encodes that a question is something ‘looking for’ an expression of category X to give a question-answer pair. (For convenience, however, I continue to use the label Qu[X] rather than its categorial grammar recasting.)

The semantics to be developed below goes hand in hand with this. First, I take the meaning of the Qu in 1a to be the function represented informally in 7a, and the meaning of the Qu in 5a to be 7b (this is also the tack taken in Ginzburg & Sag 2000).

- (7) a.  $\lambda x[x \text{ left the party at midnight}]$   
 b.  $\lambda x[\text{Bozo invited } x \text{ to the party}]$

Then the semantics of a Qu-Ans is such that the function corresponding to the Qu part is applied to the meaning of the Ans part. This means that the relevant propositions that are understood in the discourses in 1 and 5 are not just the result of general inferencing strategies. Rather, this is part of the grammar: it is the grammar of the construction that combines the meanings of the two parts. Note that this means that main-clause questions do not denote sets of propositions (as in Hamblin 1973 and Karttunen 1977), nor

---

putting the meaning of the question together with the meaning of the answer, and so neither single speaker has literally uttered the proposition expressed by the full Qu-Ans. This raises some interesting and deep questions about the status of such shared propositions, but I leave these as open questions.

do they denote functions from worlds to propositions as in Groenendijk & Stokhof 1984. Rather, they are simply functions of type  $\langle x, t \rangle$  for some  $x$ . (And multiple WH-questions are of type  $\langle x, \langle y, t \rangle \rangle$ .) Here we agree with Ginzburg and Sag (2000), who also take this as the meaning of main-clause questions, and agree in part with Groenendijk and Stokhof (1989), who take this as one meaning for a main-clause question. But notice quite crucially that this in no way precludes the possibility that EMBEDDED questions have the Hamblin/Karttunen meaning (or the meaning in Groenendijk & Stokhof 1984); that meaning can be derived from 7 in a systematic way, and so main-clause questions can shift to become embedded questions. And since the evidence for the Hamblin/Karttunen semantics is based entirely on the compositional contribution of embedded questions, giving main-clause questions the meaning in 7 has no effect on theories of embedded questions. Moreover, there is clearly some extra bit of ‘meaning’ in main-clause questions that we can think of as the illocutionary force of a question. For surely the question *Who came?* and the VP *came* are different creatures, and so questions do not have the same meaning in the broadest sense as that of VPs. I have no particular theory in mind here of how to concretely encode the illocutionary force, but I assume that in addition to the normal sense of ‘meaning’, there is an illocutionary force operator associated with a question that makes it a request for information.<sup>12</sup>

The remarks above need slight elaboration, for Groenendijk and Stokhof (1984) note that generalized quantifiers such as the expressions in 8b–e are all also perfectly good answers to a question like 8a.

- (8) a. Q: Who left the party at midnight?  
 b. A: Everyone.  
 c. A: No one.  
 d. A: No one except Claribel.  
 e. A: Claribel or Bozo.

<sup>12</sup> Groenendijk and Stokhof (1989) argue that while 7a is one meaning associated with the question *Who left the party at midnight?*, it cannot be the only meaning, and they propose a type-shift rule to derive one meaning from the other. While this is compatible with the proposal here, it should also be noted that their arguments against having 7a as the only meaning do not seem very strong. The one with the greatest force at first glance concerns coordination of questions, as in (i).

- (i) Who danced and who drank wine?

Incidentally, under any analysis there is presumably a reading of (i) in which it is a multiple WH-question (and hence has a meaning of type  $\langle e, \langle e, t \rangle \rangle$ ), but the reading of interest here is the one in which this ‘feels like’ a coordination of two questions. This is brought out with the intonation suggested by the punctuation in (ii).

- (ii) Who danced? And who drank wine?

But as Groenendijk and Stokhof (1984) point out, (i) and (ii) (under the two-questions reading) do not mean the same thing as does a question like (iii), which has conjoined VPs. Yet if *Who danced?* and *Who drank wine?* have the same meanings as *danced* and *drank wine*, respectively, this difference is unexplained.

- (iii) Who danced and drank wine?

But it is not clear that (i) and (ii) involve ordinary coordination (unlike the case of (iii)); Krifka (2001) analyzes these as instances of ‘speech act coordination’. Recall that we are assuming that questions come with an illocutionary force operator that distinguishes them from VPs, and we might suppose that this makes them immune from ordinary coordination, preventing (i) and (ii) from having a meaning like that of (iii). As evidence for this position, notice that ‘coordinated’ questions of this type do not show the full range of possible coordination behaviors.

- (iv) \*Both who danced and who drank wine?

Similarly, while questions can (albeit somewhat marginally) be connected with *or*, they do not occur with *either ... or*.

- (v) a. ?Who danced? Or, who drank wine?  
 b. \*Either who danced or who drank wine?

This fact will play a crucial role later. Given the standard assumption that expressions like those in 8b–d are of type  $\langle\langle e, t \rangle, t\rangle$ , then in this case the semantics is put together by having the answer take the question semantics in 7 as argument. This can be done via a type-driven principle (the semantics of the Qu-Ans is put together by having the function corresponding to Qu take the Ans denotation as argument or vice versa, whichever matches the types). Note too that the fact that generalized quantifiers can be answers has a small consequence if this is embedded in a categorial grammar syntax, according to which *everyone*, for example, would be of category  $S/(S/NP)$ . In this theory, the syntax of Qu-Ans would require that the two parts be either  $Qu[X]$  and  $X$ , or else be  $Qu[X]$  and  $S/(S/X)$ .

While the above analysis is largely taken from Groenendijk and Stokhof (1984), I depart from them in two key ways. First, they build exhaustification into the semantics that combines the question with the answer. In other words, their rule is more complex, and it ensures that the proposition resulting from the dialogue in 1 is that Claribel and only Claribel left the party at midnight. But, following Schulz and von Stechow (2006), among many others, I assume that the listener's conclusion that only Claribel left the party at midnight is a more general pragmatic fact and is not forced by the semantics. Second, Groenendijk and Stokhof (1984) consider a full sentence like *Claribel left the party at midnight* to be a full-blown answer as well as the short answer. Crucially, I argue that *Claribel* and *Claribel left the party at midnight* have a different status in the dialogue in 1. The short answer is a true linguistic answer in the purely technical sense used here of a Qu-Ans construction. The full sentence is a REPLY: it obviously supplies relevant information to the listener (in this case, it supplies exactly the same information), but it is not a genuine 'answer' in the technical sense above. To forestall any confusion about the grammatical category of an 'answer' and the everyday use of this term, I refer to the latter as a 'reply' and reserve 'answer' for the technical sense of this notion in the Qu-Ans theory. Note, though, that the grammar itself has no category label 'answer'; the category labels of 'answers' are things like NP, PP, and so forth. So by 'answer' is meant any expression that forms a unit with a Qu (which is a grammatical category) to give a Qu-Ans. Hence, the technical notion of an 'answer' is parasitic on the notion of a Qu-Ans.

There is one further issue that I note without solution. This is the question of exactly how to state the syntax of Qu-Ans, in that the question and answer need not actually be adjacent in the discourse; there can be various material intervening, as in the discourse in 9.

- (9) a. Q: Who left the party at midnight? Do you know?  
 b. A: Yeah, um ... Bill.

Exactly how to characterize the relation between the Qu and the Ans undoubtedly requires a fuller theory of the structure of discourse.

**3.3. YES/NO QUESTIONS AND ALTERNATIVE QUESTIONS.** The above centered on WH-question-answer pairs. An obvious question to ask is whether the analysis extends in a natural way to yes/no questions. The answer (no pun intended) is 'Yes'. Here I again follow Groenendijk and Stokhof (1984), whose proposal is that the semantics of a main-clause question such as *Did John swim?* is the function  $\lambda o[o(\llbracket \text{John swim} \rrbracket)]$ , where  $o$  ranges over two functions of type  $\langle t, t \rangle$  (extensionalizing):  $\lambda p[p]$  and  $\lambda p[\sim p]$ . That is,  $o$  has two values: the identity function on propositions and negation, or alternatively  $\llbracket \text{yes} \rrbracket$  and  $\llbracket \text{no} \rrbracket$ . Ultimately we need to fold in intensions, in part to accommodate modifier meanings such as  $\llbracket \text{possibly} \rrbracket$ ,  $\llbracket \text{probably} \rrbracket$ ,  $\llbracket \text{certainly} \rrbracket$ , and so forth.

Then the actual linguistic answers (in our technical sense) to *Did John swim?* are *yes*, *no*, *possibly*, and so forth. As with the case of WH-questions, a full sentence like *John swam* or *John did (indeed) swim* is simply a long reply. Alternative questions are also unproblematic on this approach. An alternative question such as *Did she decide to take phonology, or syntax?* has as its meaning  $\lambda_{x_e \{ \text{phonology, syntax} \}} [ \text{she decided to take } x ]$ . In that case, the expression *phonology* is an appropriate ‘answer’ to form a Qu-Ans pair.

**3.4. MAYBE JOHN.** One might wonder about short answers like *maybe John*, *probably Bill*, *possibly Tom*, and so forth. It is common to think of *maybe*, *probably*, *possibly* as sentence operators, and so at first glance these short answers seem to provide clear evidence for SLM. But in fact we can treat these fragment answers as generalized quantifiers. Indeed, (as is also the case with *not John*) these phrases coordinate with ordinary NPs and with generalized quantifiers, which is not expected if *not*, *maybe*, *possibly*, and so forth were only sentence modifiers.

- (10) a. Bill and/but not John left.  
 b. Bill and maybe/probably/possibly John left.  
 (11) a. Every girl and/but not John left.  
 b. Every girl and maybe/probably/possibly John left.

Spelling out the full semantics of an expression like *maybe John* requires a modal semantics and so is not done here, but once we have a modal semantics it is clear that  $\llbracket \text{maybe John} \rrbracket$  can denote the set of properties that John might have. (We will be looking at the case of *not John* in greater detail in §4.1; see the discussion surrounding 24.) Interestingly, other types of adverbs, such as subject-oriented adverbs, cannot be a piece of the short answer, and—as the remarks above would lead us to expect—these also do not occur as part of a complex generalized quantifier.<sup>13</sup>

- (12) a. Q: Who left?  
 b. A: \*Carefully John. (compare to *Carefully, John left.*)  
 (13) \*Bill and carefully John left.

**4. ADVANTAGES OF QU-ANS OVER SLM.** Many arguments have been given for the view that there is silent linguistic material in the position of the ellipsis site; I postpone discussing these until §5. First I present three new arguments for Qu-Ans over SLM: these center on facts that follow immediately under the Qu-Ans analysis but are problematic under SLM. The first two are related (§§4.1 and 4.2) and concern the semantics of answers in conjunction with the questions. To preview these briefly: the proposition inferred from the combination of question and fragment answer is not always the same as that expressed by the corresponding long reply. This is unsurprising under Qu-Ans, for here the question itself also contributes to the semantics of the Qu-Ans proposition. A simplistic view of SLM has nothing to say about this (the proposition conveyed should just be the same as that of the long ‘reply’), and so one might try a more sophisticated version that requires some connection between the silencing of linguistic material in the reply and the semantics of the question itself. In other words, something will be needed to ensure that silencing is allowed only in the case of a ‘reply’ that is a genuine answer to the question. Not only does this remove any advantage that SLM might claim in virtue of not needing a notion of an answer (or a Qu-Ans pair), but—more seriously—we will see that in fact there is no obvious way to define the requisite notion of an ‘answer’. Hence it is not clear that SLM can account for the facts. The third argu-

<sup>13</sup> I thank Geoffrey Pullum for this observation.

ment against SLM (§4.4) centers on CASE MARKING, and it is especially telling in that case marking has classically been taken as an argument FOR SLM. But a closer look reveals that SLM—unless heavily supplemented with additional principles—actually does not account for the facts, while Qu-Ans immediately does.

**4.1. PRESUPPOSITIONS CONTRIBUTED BY THE WH-WORD.** Consider the questions below, followed by their short answers (the (b) examples), and contrast these with the long replies (the (c) examples).

- (14) a. Q: Which mathematics professor left the party at midnight?  
 b. A: Jill.  
 c. A: Jill left the party at midnight.
- (15) a. Q: Which students who had come to the party without costumes were awarded prizes (anyway)?  
 b. A: Claribel and Bozo.  
 c. A: Claribel and Bozo were awarded prizes.

In 14b, the responder is committed to the belief that Jill is a mathematics professor; in 15b the commitment is to the fact that Claribel and Bozo are students and that they came to the party without costumes. No such commitment holds if the responder uses the fuller replies in 14c and 15c. Indeed, quite the contrary: 14c at least strongly suggests that the responder is not certain that Jill is a mathematics professor, and likewise 15c suggests that the responder is not certain that Claribel and Bozo are students who came to the party without costumes. They are thus most natural when said with the intonational pattern known in the literature as the FR (fall-rise) pattern (see e.g. Ward & Hirschberg 1985) and also when preceded by *well* (uttered with that same pattern). Without that intonation (and perhaps also without *well*), the long replies actually seem unnatural; it is difficult to figure out why a normal responder would use these. I thus call the long replies with the appropriate intonation BEST-I-CAN-DO replies. The FR intonational pattern is found in a more general set of cases (again see Ward & Hirschberg 1985); here it is used because the responder is giving information whose direct relevance to the question is unclear. For convenience, I also call the responder's belief regarding Jill in the answer in 14b a 'presupposition' and refer to this as the 'presupposition contributed by the WH-word'. This glosses over some thorny issues about whether this terminology is really appropriate, but I believe that it is harmless for the purposes at hand.

The argument to be developed below rests on two observations. The first and most important one is that the short answer CANNOT have the best-I-can-do reading in the sense of the responder being uncertain that Jill is a mathematics professor—the short answer commits the responder to the belief that Jill is a mathematics professor. (It might be able to have a best-I-can-do reading and intonation for other reasons, which we return to below.) The second observation is that the long reply is at best quite odd without the best-I-can-do reading and intonation. (There are long replies without this requirement; these are also discussed below.)

HOW THE FACTS FOLLOW UNDER QU-ANS. The first observation follows immediately in the Qu-Ans account. We assume that the WH-word contributes to the semantics of the question, and so the question denotes a partial function defined only for mathematics professors.

- (16)  $\lambda x_{xe[[\text{mathematics professor}]]} [x \text{ left the party at midnight}]$

This then can combine with  $[[\text{Jill}]]$  to give a Qu-Ans proposition only if Jill is indeed a mathematics professor.

The fact that 14c most naturally has the best-I-can-do reading is also unsurprising given some fairly innocent pragmatic assumptions. In the view here, it is not a true ‘answer’ but a ‘reply’. Because it is not a linguistic answer and also because it is longer than needed, we can assume that a responder would opt for this only if there is some reason to do so. In other words, the short answer both is a better form (we know that in general there is a penalty for repeated material) and it makes a better contribution in terms of providing the questioner with the information that is wanted. This is because it is a true answer and hooks to the semantics of the question, so there is no doubt that 14b as part of a Qu-Ans provides the listener with the information about which mathematics professor left the party at midnight. But—under the Qu-Ans analysis—the long reply in 14c does not have the same status; it is just a sentence in its own right whose meaning is a proposition. Since the responder gives this reply, the questioner can assume that the proposition in 14c has some relevance to the discourse. But the semantics of the long reply itself does not enforce any tight connection with the question. Hence, the original questioner can also conclude that there is a good reason that a responder opted for this form as opposed to the true ‘answer’. And since nothing about the meaning of this sentence by itself assures that Jill is indeed a mathematician, an obvious reason for using the long form would be the very fact that it does not commit to Jill’s mathematician status. Hence the best-I-can-do inference.

There are two conceivable responses one might make to the above line of argument. The first centers on the explanation for why the long reply generally has a best-I-can-do interpretation, for there are instances of long replies that are quite naturally understood without this. Many speakers (though not all) seem to be fine with a VPE answer like 17b as indicating that the answerer is committed to Jill as mathematics professor.

- (17) a. Q: Which mathematics professor left the party at midnight?  
 b. A: Jill did.

A potential explanation for the fact that 17b is quite natural without a best-I-can-do reading stems from a tension between two principles. On the one hand, the short answer is a better form because it is the real answer. On the other hand, there also seems to be a prohibition against ‘curtness’ in actual conversation. (I am grateful to Hugh Rabagliati for suggesting this explanation.) Thus there is a competition between the true answer *Jill* and the VPE version in 17b, so a listener can assume that 17b is used to comply with the ‘Don’t be curt’ prohibition rather than assume that it is because s/he is unsure of Jill as math professor. But the full long reply in 14c has no advantage over either the short answer or 17b, so it strongly suggests that there must be some other reason for its use. Note, incidentally, that one might then be tempted to say that the only reason that the full long reply in 14c is odd is because it competes with 17b (with VPE), and that the competition with the short answer is irrelevant. But this fails to explain why it is that 14c has the particular best-I-can-do interpretation that it does. The fact that it strongly suggests that the responder is not sure that Jill is a mathematics professor follows only via comparison with the short answer (not via a competition with 17b).

There is at least one other case where a long reply does not favor a best-I-can-do reading for the simple reason that the competition with the short answer does not exist: this centers (for many speakers) on multiple WH-questions. Thus consider 18 in a *Fahrenheit 451*-like context where—knowing that the House of Representatives was about to pass a law burning all books—each member of the Senate decided to memorize one book.

- (18) Q: Which New England senator memorized which book?

For many speakers, the syntax of English is such that there is no short answer. (Thus this can never actually enter into a full-fledged Qu-Ans.) This is because pairs of NPs are not well-formed expressions in their own right, so 18 cannot be answered by 19 (the relevant characters were all New England senators in 2010).

(19) A: \*Sanders—*Das Kapital*, Lieberman—*War and Peace*, and Kerry—*Moby Dick*.

It should be noted that I have found a few speakers who do allow 19, but it is at least somewhat degraded for many. Hence, the full long reply is predicted to be fine without a best-I-can-do reading (at least for those who find 19 odd). And indeed this is the case; 20 is impeccable if the speaker believes that Sanders, Lieberman, and Kerry are all New England senators.

(20) A: Sanders memorized *Das Kapital*, Lieberman memorized *War and Peace*, and Kerry memorized *Moby Dick*.

The second objection to the line of reasoning above centers on the first observation: that the short answer cannot have the best-I-can-do reading. I have claimed that this is because the semantics of the question is crucial to the semantics of the Qu-Ans, so the restriction on the domain of the WH-word crucially figures into the semantics of the Qu-Ans because the semantics of the question and of the fragment jointly determine the Qu-Ans proposition. But one might think that there is a different reason for this: one might suspect that the only reason that the short answer cannot (in general) have the relevant best-I-can-do reading is simply that it is not long enough to host the necessary intonation. But this can be shown to be incorrect on two grounds. First, short answers ALWAYS carry the presupposition of the head even when they are definitely long enough to host the relevant prosody. Second, in a limited set of circumstances, even very short answers CAN have a best-I-can-do reading and prosody, but this is for other reasons (except in one special case discussed below).

As to the first point, consider the dialogue in 21.

(21) Q: What silly thing that Donald heard from Sarah did he actually believe?

A: That someone put two phony birth announcements in Hawaiian newspapers to conceal the birth of a baby in Kenya.

Even if one tries to put the right prosody on this, the answerer is committed to the view that the phony birth announcement scenario is something that Donald heard from Sarah (and is silly). In fact, it seems odd to put a *well* in front and give it the right prosodic contour, although one can find some other reason why this might be possible, as is shown for a simpler case below. The same point is made by the dialogue in 22; the short answer NP is not in any obvious sense too short to host the relevant prosody, but here too the responder is committed to the belief that the person in question is a mathematics professor.

(22) Q: Which mathematics professor left the party at midnight?

A: The guy I was introduced to this afternoon at lunch.

As to the second point, there are indeed cases where one can construct a scenario in which a short answer does have a best-I-can-do reading and intonation. But this is not because the responder is unsure of whether the presupposition of the head is met. For example, Ward and Hirschberg (1985) point out that the FR intonation is possible in cases where the responder is not sure if the Qu-Ans proposition provides the complete information that the questioner is looking for, as in the dialogue in 23.

(23) Q: Which mathematics professors left the party at midnight?

A: Well, Jill ... I'm not sure if anyone else did.

Here the responder is still committed to the belief that Jill is indeed a mathematics professor; the best-I-can-do part arises because the answerer assumes that the questioner wants a complete answer, and the perhaps partial answer is the best she can do.

Similarly, a best-I-can-do reading and FR intonation is easily found in 24.

(24) Q: Which mathematics professor left the party at midnight?

A: Well, not Jill.

This case is particularly interesting, for several reasons. First, the distribution of a generalized quantifier like *not Jill* is puzzling (under any account); it can occur alone in fragments and in conjunctions like *Jeff but not Jill left the party at midnight* but not alone in argument position as in *\*Not Jill left the party at midnight*. One might at first glance think that this itself argues against the SLM account of the answer in 24a, but in fact both theories are going to need a fuller account of the distribution of generalized quantifiers like this, so I do not claim any victory here. I do, however, assume that its meaning is the set of properties that Jill does not have (i.e. those that do not map Jill to 1). Second, the most natural understanding of the dialogue in 20 (with or without the FR intonation on *not Jill*) is that the responder is committed to the belief that Jill is a mathematics professor. Under that understanding, the FR intonation (if present) is due to the fact that the responder knows that this is not the kind of reply the questioner was hoping for—it does not specify who actually did leave the party (nor is it a complete listing of the leavers). Nonetheless, it has—as predicted—the Qu-Ans semantics since it formally qualifies as a Qu-Ans, so the proposition that Jill did not leave the party at midnight (combined with the commitment that she is a mathematics professor) follows from the dialogue.

It is important to note that the fact that *not Jill* can have the FR intonation in virtue of its not giving complete enough information means that any theory needs a characterization of what constitutes a ‘fully satisfactory response’. Such a response is something that (possibly by itself or as part of a Qu-Ans) gives the full and complete information the questioner is presumably hoping to learn. From that we can also derive the complementary notion of a best-I-can-do response: it is something that gives potentially relevant information but is not a fully satisfactory response. This notion of a fully satisfactory response is presumably a fact about questioner expectations and general principles of cooperative dialogue. But it is important to note that a fully satisfactory response in this sense is not always same as the proposition that corresponds to the Qu-Ans, as 24 shows. To recast the discussion in theory-neutral terms, 24 shows us that the notion of a fully satisfactory response (which 24 is not) is distinct from the notion needed to characterize WHAT CAN STAND ALONE AS A SHORT ANSWER. The importance of this for the discussion here is that it is argued below that SLM ultimately needs some notion of an ‘answer’ in order to predict the distribution of Shh (or of FP). One might therefore be tempted to claim that no extra burden is placed on SLM, for every theory needs to characterize the notion of a fully satisfactory response. But that conclusion is incorrect: the theory-neutral notion of a fully satisfactory response is simply not the same notion of ‘answer’ that will be needed to account for the distribution of fragment answers. Indeed, this is exactly the moral to be drawn from the dialogue in 24 (on its very natural understanding in which the responder does believe that Jill is a mathematics professor but still uses *Well, not Jill* with FR intonation). Here the fragment answer is fine; nonetheless, it is not a fully satisfactory response.

There is a further interesting point about 24. Unlike all of the other short answers we have seen so far, it seems marginally possible to get the FR intonation on this if the responder is NOT committed to the assumption that Jill is a math professor. In this case,

this is a best-I-can-do answer precisely because the responder does not accept the presupposition of the WH-expression. I have encountered variability on how tolerant speakers are of this reading, but it helps to bring this out by elaborating the answer as follows: *Well, not Jill—I don't think she's a mathematics professor*. But this is not surprising, given the well-known fact that while negation in general negates assertions and preserves what I loosely call the presupposition, it can also be used (perhaps marginally) in cases where the presupposition is not met. Note that this is no different from other instances of negation with presupposition violation (as in *Well, Jill is not the mathematics professor who left—she isn't a math professor*, where here too is found FR intonation).

In sum, then, the Qu-Ans analysis accounts for the following complex array of facts: (i) except in the *not Jill* cases, the fragment answer always preserves the presupposition of the WH-expression; (ii) in general, the long reply is most naturally understood as a best-I-can-do response (preferring FR intonation) and thus does not preserve this presupposition; (iii) but some cases of long replies are quite perfectly fine without the best-I-can-do reading (and intonation) because the competition with the fragment (true) answer is absent (as in multiple WH-questions) or is mitigated by other factors; and (iv) a short answer like *not Jill* may for some speakers allow for the relevant best-I-can-do reading for the reason that, in general, presuppositions can sometimes not project under negation. I now turn to SLM to see if it accounts for the facts.

**4.2. THE PROBLEM FOR SLM.** What might the SLM view say about these facts? As to the observation that the long reply generally requires a best-I-can-do reading (point (ii) above), essentially the story told above can be imported directly into SLM as long as it can also provide an account of the fact that the short answer preserves the presupposition of the WH-expression. The point in (iii) above will also carry over, provided, again, that there is some way to account for the observation in (i). In other words, assume under SLM that long answers are worse forms because there is a penalty for repetition (a fairly well-documented assumption). Then as long as there is an account of why it is that the short answer has the relevant presupposition, we can explain why the long answer generally does not. The Gricean story told above—with some modification—could carry over directly to SLM, once there is an account of why it is that the short answer necessarily carries the presupposition of the WH-expression.

But indeed, therein lies the rub: the difficulty comes in ensuring that the short answer commits the responder to the presupposition contributed by the head. Clearly some connection has to be enforced between the question and the short answer. Hence, an advocate of SLM might hope that there is some way to restrict Shh to appear only in sentences that are actually 'answers' to the relevant question. Initially one might hope that this is a consequence of independently motivated conditions on the relationship between focus and ellipsis (see e.g. Rooth 1992); evidence is presented in the next section that focus considerations are not enough to predict the distribution of Shh (or FP). Hence it appears that the grammar must simply incorporate some notion of 'answer' into the question and license Shh (or FP) only in sentences that are answers. But this has problems. First, one must then give up any hope that the conditions on Shh derive from more general conditions regarding ellipsis. More seriously, we will see that there is no obvious way to define the right notion of 'answer' as to predict the distribution of short answers.

**DERIVING THIS FROM FOCUS SEMANTICS?** First, consider the semantics of a question like 14a under the question semantics of Hamblin 1973. Here 14a denotes a set of propositions of the form {Jill left the party at midnight, Jeff left the party at midnight,

Michael left the party at midnight, ... }, where in each case the relevant party-leaver is a mathematician. If we can ensure that the proposition corresponding to elided answers is a member of this set, we would successfully capture the fact that the short answer bears the presupposition introduced by the WH-expression.

One might, then, hope to derive this from more general principles governing ellipsis using the theory of VPE from Rooth 1992, which assumes the theory of focus in Rooth 1985, where expressions have a regular semantic value and a focus value. Consider a sentence with ‘focal’ or ‘contrastive’ stress such as *LUCY ordered lobster* (with prosodic prominence on *Lucy*). Here, the regular value is just the ordinary meaning of the sentence (a proposition), and the focus value is a set of alternative propositions such as {Sally ordered lobster, Mary ordered lobster, Julie ordered lobster, ... }. Given this, Rooth proposed that VPE is subject both to some sort of identity condition and to the additional condition in 25.

- (25) An elided VP must be contained within some constituent  $C_{ELL}$  that has a focused constituent within it, and the antecedent VP must be contained within a constituent  $C_{ANT}$  such that the ordinary value of  $C_{ANT}$  must be a member of the focus value of  $C_{ELL}$ .

It would be reasonable to suspect that this extends directly to the case of ellipsis in short answers. Using Merchant’s mechanics, a possible approach would be to assume that a full FP contains focus—presumably on the fronted part, which is the remnant in the short answers—and that there must therefore be some expression (i.e. the question itself) that is a member of the focus value of the FP. Indeed, Merchant points out that this possibility is the very reason he named the expression ‘FP’.<sup>14</sup>

But this will not do. In the first place, the question is not a member of the focus value of the FP. In fact, under the Hamblin view of the semantics of questions, the regular value of the question is identical to the focus value of the FP. One could of course extend Rooth’s condition to say that the larger constituent  $C_{ANT}$  (in this case the question) must either be a member of OR be identical to the focus value of the FP (i.e.  $C_{ELL}$ ). But this undermines the claim that one can derive the ellipsis condition here from an independent condition on focus with respect to ellipsis: the focus condition on ellipsis is revised to a disjunctive condition, and the second disjunct has no independent motivation.

But aside from this objection, this still does not work. For, as noted earlier, all sorts of generalized quantifiers can be good short answers. Yet the propositions corresponding to the corresponding long replies (e.g. *No one but Jill left the party at midnight*) are not members of the Hamblin question.

- (26) a. Q: Who left the party at midnight?  
 b. A: No one but Jill.  
 b. A: Two mathematics professors, three deans, and every clown.  
 c. A: Claribel or Bozo  
 d. A: Every clown.  
 e. A: No one.

<sup>14</sup> It should be noted that FPs in Merchant’s account obligatorily contain Shh in the head position. Put differently, the silencing must take place. For although we do get topicalized sentences with fronted material, as in (ib), these are entirely inappropriate as answers to questions, as shown in the full dialogue in (i).

- (i) a. Q: Which clown did you invite?  
 b. A: #Bozo, I invited.

Presumably, the fronting of *Bozo* in other cases would involve fronting to the Spec position some other node such as perhaps TopP, while in a simple answer like *Bozo* it fronts to Spec of FP.

Hence, the Hamblin question neither is in nor is the focus value of, for example, *NO ONE BUT JILL left the party at midnight* (with focus on *no one but Jill*).<sup>15</sup>

Finally, there is additional evidence that the set of environments allowing focus and the set allowing short answers are not the same. We find cases where a generalized quantifier can be focused in a full sentence, but where the corresponding fragment is nonetheless impossible—even when an identity condition on ellipsis would be met. Suppose, for example, that Sheila and Bill are discussing how hard Professor Carberry's ceramics courses always are and wondering whether anyone did well this year on the final. Crucially assume a discourse in which there is no immediate explicit question of the form *Who got an A the exam?*, although in the context set up here such a question is implicit. Moreover, Carberry is eavesdropping and steps into the conversation after Sheila's utterance in 27a. In such a context, the dialogue in 27 is impeccable.

- (27) a. Sheila: Oh, I just remembered. BOZO got an A on the exam.  
 b. Carberry (butting in): Wrong. No one but CLARIBEL got an A.<sup>16</sup>

But the short answer in 28b is impossible in this same context.

- (28) a. Sheila: Oh, I just remembered. BOZO got an A on the exam.  
 b. Carberry (butting in): Wrong. \*No one but CLARIBEL.

Notice that we cannot pin the unacceptability of 28b on the fact that identity is not met—obviously there is a linguistic antecedent of the form *got an A on the exam* (or *t got an A on the exam*). So it is obviously not enough to license ellipsis by a combination of identity and focus. As such, it is difficult to imagine any condition on focus alone that

<sup>15</sup> A referee raises an important question: could not the Hamblin semantics simply be revised to say that the question denotation is the set of all propositions of the form *X left the party at midnight* for X any generalized quantifier? If so, then  $\llbracket$ No one but Jill left the party at midnight $\rrbracket$ ,  $\llbracket$ No one left the party at midnight $\rrbracket$ , and so forth would all be part of the (revised) Hamblin question. While this might be possible, it undermines the independent motivation for the Hamblin semantics and would cause serious complications for the analysis of embedded questions. As discussed by Beck and Rullman (1999), a sentence like (i) does not commit to a strongly exhaustive reading but can have what is known rather as a weakly exhaustive reading.

(i) Becky knows who came to the party.

The weakly exhaustive reading is one in which for all individuals x, if x came to the party, then Becky knows that x came to the party. However, if y did not come to the party, she need not know that. (For extensive discussion of this semantics, see Beck & Rullman 1999.) One can easily account for this by defining the meaning of question-embedding *know* (call that *know<sub>Q</sub>*) in terms of proposition-embedding *know* (call that *know<sub>p</sub>*). Thus  $\llbracket$ know<sub>Q</sub> $\rrbracket$  is such that it combines with its object (the Hamblin question) and returns true if and only if for all true p in the Hamblin question, the subject knows p. Other techniques are possible here as well, but the result in all cases would be that if the object of question-embedding *know* included all propositions about generalized quantifiers, there would be no way to get the weak exhaustive reading. Rather, only the strong exhaustive reading would be possible; I leave this to the reader to verify.

Accounting for the 'mention-some' reading found in (ii) would also be problematic.

(ii) John knows where to get gas.

Under the actual Hamblin semantics, this can be accounted for by shifting  $\llbracket$ know<sub>Q</sub> $\rrbracket$  to require just one true proposition. But surely if John knows that one can get gas at either the local Cumberland Farms or the gas station on Thayer and he knows nothing else, then (ii) is not true. But if the shifted *know* merely required there to be one true proposition in John's 'know'-set, and the Hamblin semantics were recast in the way suggested above, then (ii) should be true under this scenario.

<sup>16</sup> I am assuming that the stress on *Claribel* is the manifestation of phrase-level focus (the entire NP *no one but Claribel* is focused rather than just *Claribel*). This is because what is at issue here is who passed the exam, not who is such that no one but them passed the exam. This does raise some interesting questions about computing focus alternatives: if the generalized quantifier  $\llbracket$ no one but Bozo $\rrbracket$  can be a member of the focus value of *CLARIBEL* (and  $\llbracket$ Claribel $\rrbracket$  can be a member of the focus value of *no one but BOZO*), then the computation of focus alternatives involves generalized quantifiers in some cases as well as individuals. But this is orthogonal to the points here.

would allow an elliptical answer like 26b and not also allow ellipsis in 28b. Yet ellipsis is impossible: *(t) got an A* cannot be ‘silenced’. I should note that there are some cases of fragments as answers to implicit questions, and the conditions for these (as well as their analysis) are left open here (see online appendix B). Thus *Wrong, only CLARIBEL* seems marginally possible as a follow-up to 27a; just why this is better than 28b is not explored here. But for the present purposes, it is sufficient to note that certain expressions can be short answers to questions, but cannot serve as fragments even when they are focused and when an identity condition would be met. This then suggests that under an SLM account of short answers, the grammar really will need some way to allow ellipsis in short answers just in case these are answers to questions. Hence, some connection between the question and the answer must be enforced. Note that the points in this section make good on a promise given when the term ‘Shh’ was introduced in §3.1: the conditions on silencing (or on the distribution of Shh or ‘FP’) do not follow from more general conditions on silencing.

DEFINING ‘ANSWER’ AS A CONDITION FOR SHH? At this point, an advocate of SLM might hope that—armed with an appropriate notion of an ‘answer’—the fact that the fragment answer preserves the presupposition of the head can be accounted for. Assume simply that an FP must be an ‘answer’ to a question. Since Shh is the head of FP (and thus occurs nowhere else), it then follows that the silencing licensed by Shh will only be found in ‘answers’. And perhaps the relevant notion of ‘answer’ is such that it inherits the presupposition of the head. From all of this, it would follow that the phonetic form of an answer in which part is silent will carry this presupposition with it.

But how can we define the appropriate notion of an answer? Starting again from the Hamblin semantics, one might hope that an answer to a question is something in the Hamblin set denoted by the question. But once again, generalized quantifiers as short answers are not accommodated under this view, for the proposition expressed by 29b is not a member of the Hamblin question corresponding to 29a.

- (29) a. Q: Who left the party at midnight?  
 b. A: No one but Jill left the party at midnight.

Similar remarks apply to all of the short answers shown earlier in 26.

There is one more tack one might try here. The point of departure for this strategy is the partition semantics of Groenendijk and Stokhof (1984) for questions. According to this, a question like 30 partitions the space of possible worlds into ways the world could be.

- (30) Q: Which deans left the party at midnight?

To exemplify, let Claribel, Bozo, and Tinkerbelle all be deans at Clown Tech. Then 30 partitions the world into eight subsets or ‘cells’. (I ignore the contribution of singular and plural morphology here and for convenience treat *which dean* and *which deans* alike.) The cells then are as follows: (i) all worlds in which Claribel, Bozo, and Tinkerbelle left; (ii) the worlds in which only Claribel and Bozo left; (iii) those in which Bozo and Tinkerbelle left; and so forth.

Suppose then that something ensures that an FP must remove one or more of the cells in this partition. It is thus a ‘partial answer’—it does not need to zero in on the exact cell but it does need to remove some. In a question like *Which deans left the party at midnight?*, a simple answer like *Claribel* [~~*t left the party at midnight*~~] succeeds in removing a partition only if Claribel is indeed a dean. In the scenario above, an answer like *Grumpy* [~~*t left the party at midnight*~~] has nothing to do with the relevant partition, and so *Grumpy* cannot be a short answer. A generalized quantifier answer like *Claribel or*

*Bozo* [~~*t left the party at midnight*~~] also removes at least one cell, and so it follows that this can be an FP containing *Shh* as its head. In other words, the hypothesis on the table here is that for the purposes of being an FP (and containing *Shh*), an answer is any sentence conveying a proposition that removes at least one cell in the Groenendijk and Stokhof partition.

If this is what SLM needs in order to predict the right semantics for the category FP (and hence the distribution of *Shh*), then one might already be highly suspicious of the claim that there is any obvious simplicity to the SLM view. FP is the name of a category whose semantics is such that it connects to a question in some way. Specifically, the proposition that is the meaning of FP must remove one of the cells in the Groenendijk and Stokhof partition (or, put differently, it must partially answer the question). It is not at all clear how to syntactically formalize this: the grammar must ensure category label FP is allowed only if the proposition denoted by the FP connects to a question in just this way.

But even this will not get the full range of facts. Note first that this makes a different prediction from Qu-Ans for the case of a fragment like (*Well*), *not Claribel* as an answer to 30. Under the SLM solution currently under discussion, *not Claribel* can serve as a short answer only if *Claribel* is a dean, for it is only under that situation that a cell in the Groenendijk & Stokhof 1984 partition for 30 is removed. Yet we have seen that (at least marginally) this can be used (with FR intonation) when the responder is not committed to *Claribel*'s deanhood; here the presupposition need not survive in the same way that negation can in general (marginally) be a 'hole' for presupposition. Moreover, there is a second case that is problematic for this version of SLM, to which we now turn.

**4.3. EXPLICITLY EXHAUSTIVE QUESTIONS.** So far, we have seen that SLM might try to account for the distribution of short answers under the hypothesis that they are allowed just when the proposition removes one cell in the Groenendijk and Stokhof partition. Aside from the fact that this seems quite stipulative, this avenue of solution does not extend to the case of short answers to explicitly exhaustive questions.

Thus, consider the question in 31a and the short answer in 31b as opposed to the long reply in 31c.

- (31) a. Q: Who all left the party at midnight?  
 b. A: Bozo, Claribel, and Jill.  
 c. A: Bozo, Claribel, and Jill left the party at midnight.

Example 31b has only the exhaustive reading: the responder is committed to the proposition that these three left the party at midnight and that only these three did.<sup>17</sup> With Shulz and van Rooij (2006), I assume that the fact that exhaustiveness is suggested in the simple case of a *who* question is a pragmatic fact and is not absolutely required. But

<sup>17</sup> Roger Schwarzschild has pointed out to me that one can get 31b as a short (partial) answer to 31a without necessarily committing to the exhaustive reading if it is given the intonation of an incomplete list.

- (i) a. Q: Who all left the party at midnight?  
 b. A: Well let's see, Jill, Bozo, Claribel ...

(This is then most naturally continued with something along the lines of *and umm ... I'm not sure who else*.) Note that this does not involve the FR contour of a true best-I-can-do answer, but rather a rising intonation pattern on each of the NPs, signaling that the utterance is not complete. Hence I do not think this shows that a true full short answer need not require exhaustification; this instead seems like a special category that is marked as an incomplete utterance (the '...' above is used here as an orthographic signal of the intonation contour that suggests that the responder signals that there should be more to come). This kind of 'I have not finished' intonation can be found in any situation involving lists. Take the case of a student asking a professor

the semantics of *who all* is such that exhaustiveness is part of the meaning of 31a (see Beck & Rullman 1999 for discussion of *who all* vs. *who*). Example 31c, the (long) reply, does not have the commitment to exhaustiveness on the part of the responder. In fact, it strongly suggests that this is perhaps a partial list and that the responder is not sure if it is complete. Like the cases discussed above, it is most natural when preceded by *well* and when given the best-I-can-do prosody.

Again, these facts follow under Qu-Ans. Let the meaning of *who all* be such that the question asks for a maximal sum, and so the answer is taken as an argument of the following.

- (32)  $\lambda X[X \text{ left the party at midnight and } \forall Y[Y \text{ left the party at midnight} \rightarrow Y \leq X]$   
(where  $X$  and  $Y$  are variables over plural individuals and  $\leq$  is the ‘part of’ relation)

Thus taking the sum of the three individuals Bozo, Claribel, and Jill as argument of this, it follows that this sum must be the maximal individual. The reason for the long reply strongly wanting a best-I-can-do reading is similar to the case in §4.1. The short answer both is a better form, and—since it combines with the meaning of the question and thus necessarily commits to exhaustification—it is also a better way to answer the question. The long reply makes no such commitment since it simply denotes a proposition. The grammar does not directly connect this to the question, and so any connection is instead a fact about cooperative discourse. The questioner will thus conclude that there is some reason the responder used the long reply; the simplest conclusion is that the responder is unsure about whether Bozo, Claribel, and Jill constitute the full set of leavers.

Can we recast this using the SLM approach? No, at least not in any obvious way. The problem again is ensuring that the short answer is possible only if exhaustification is maintained. Recall that the hypothesis ‘on the table’ for the case in §4.1 is that the short answer is possible only if the proposition that it denotes eliminates some cells in the question partition. But that does not explain why the short answer in 31b requires the exhaustive reading. For indeed, even if others left as well, this does eliminate some cells in the question partition. (Note incidentally that—as pointed out in Beck & Rullman 1999—*who* and *who all* induce the same partition.)

Of course, if the theory has no way to predict that the short answer requires exhaustification, then it also has no obvious account for why it is that the long reply suggests nonexhaustification. The story under Qu-Ans as to why the long reply strongly suggests that the responder is not committed to exhaustification of the answer rests directly on the fact that there is a competing form whose semantics does require this.

Thus the conclusion from the two cases considered in §§4.1 and 4.2 is that Qu-Ans has a ready account of the presuppositions of the short answer. It also has a fairly natural pragmatic explanation for why it is that the long reply not only does not share these presuppositions but is also most naturally understood as a best-I-can-do reply. While the second result could be directly imported into the SLM analysis, this can be done only when coupled with an account of why the short answer has the presupposition that it does. But we have seen no obvious way to ensure this; at the very least it requires

---

what all the student will have to do in order to get an A in the course, and the professor responding with what is intended as an incomplete list.

- (ii) Well let’s see, you’ll have to get an A on the next three homeworks, you’ll have to turn in a decent term paper, and uh-hh ... [new intonation pattern starts] I don’t know, whatever!

So I think it is fair to conclude that (ib) is not a full short answer, but an incomplete utterance.

some reference to the question involved. This means that SLM also will need to enforce a connection between the question and the answer. Again, this shows that the silencing in fragment answers is not a more general fact about silencing operations under a theory with SLM. But even allowing for the fact that silencing can only happen in an ‘answer’, there actually is no simple way to define the appropriate semantic relation between those answers allowing Shh and the question.

#### 4.4. CASE-MARKING CONNECTIVITY.

THE BASIC FACTS AND THEIR ACCOUNT IN BOTH THEORIES. The final argument for Qu-Ans to be considered here is one that has classically been taken as an argument FOR SLM and centers on CASE-MARKING CONNECTIVITY facts discussed in Morgan 1973 and many other works since. The effect is robust crosslinguistically: case marking on a short answer must match the case one would find had the short answer actually been surrounded by fuller linguistic material identical to the relevant material in the question.

This is true even in languages with verbs that idiosyncratically assign case to their objects. Take German, for example, where verbs vary according to whether they assign accusative or dative case, and where the short answer has the case that would be found in the longer sentence. (For further examples of this type, see Morgan 1973 and especially Merchant 2004 for a wide array of crosslinguistic data.)

(33) Q: Wem /\*Wen hat Maria geholfen?  
 who.DAT/\*who.ACC had Maria helped  
 ‘Who did Maria help?’

A: Dem /\*Den Studenten.  
 the.DAT/\*the.ACC student  
 ‘The student.’

(34) Q: \*Wem /Wen hat Maria eingestellt?  
 who.DAT/who.ACC had Maria hire  
 ‘Who did Maria hire?’

A: \*Dem /Den Studenten.  
 the.DAT/the.ACC student  
 ‘The student.’

The distribution of accusative vs. dative case in the answer follows if these answers originate before deletion as full sentences containing the same verb as is found in the question.

Note first that these case-marking facts follow straightforwardly from the Qu-Ans analysis pursued here. For, as discussed above, a Qu-Ans has both a syntax and a semantics (like any other linguistic unit), and the syntax is such that the category of the constituent analyzed as Ans must match the category of the WH-expression in the Qu. All we need assume is that the category includes case information. In fact, this is exactly what Groenendijk and Stokhof (1984) and Ginzburg and Sag (2000) proposed—both sets of authors suggested that a Qu-Ans requires case matching between the category of the WH- and the category of the answer. I call this the ‘direct matching’ theory.<sup>18</sup>

<sup>18</sup> I borrow this term from Sag and Nykiel (2011), who use it with respect to case matching in sluicing. Their work extends the discussion of short answers from Jacobson 2009 to the case of sluicing.

Notice that the discussion here (in either theory) relies on the assumption that case marking is idiosyncratic and cannot be predicted by the semantics of the relevant verbs. Were there a predictable meaning for the verbs that assign dative case vs. those that assign accusative, we could also assign a slightly different meaning to dative NPs and accusative NPs. A verb like *helfen*, for example, would be defined only for those things in the set denoted by dative-marked NPs. But if this line could be maintained, the semantics would take care of the

How is the matching traditionally accounted for in SLM? At first glance this also appears to be straightforward. In the German case above, for example, matching is required not between the WH-category and the answer, but rather what matches is the rest of the question and the elided material in the answer. Both contain an instance of the verb *helfen*, and this assigns dative case to its object. Thus the case-matching condition follows straightforwardly. But note a hidden assumption here: that the requisite identity condition is FORMAL identity, for it is formal identity which ensures that the same verb (*helfen*) occurs in the elided clause. But we now turn to evidence that the requisite identity condition for deletion under SLM actually cannot be formal identity: it must be semantic identity. We return then to show that given this, case matching is actually not accounted for by SLM.

FORMAL IDENTITY CANNOT BE REQUIRED. Thus we now consider whether the identity condition for ellipsis (under the SLM view) should be formal (syntactic). This question has been much discussed throughout the history of work on ellipsis, and space precludes a full review. Part of the reason why it remains open is that for some of the cases that appear to show that the requisite identity condition is semantic and not formal (e.g. interactions with so-called ‘principle C’ effects), various strategies have been proposed to maintain that the identity condition really is formal (e.g. Fiengo and May’s (1994) ‘vehicle change’ solution). Even leaving such cases aside, however, there is strong evidence that the condition cannot be one of formal identity—it must be identity of meaning. And this evidence concerns the situation with indexicals, a point made with respect to VPE at least as early as Hankamer & Sag 1984.

Importing Hankamer and Sag’s observations into the short answer setting, take the following dialogue.

- (35) A: Who did you invite?  
B: Claribel.

To simplify things, imagine a third person C overhearing this. Notice that C understands B’s answer not to convey 36a, but rather 36b.

- (36) a. A invited Claribel.  
b. B invited Claribel.

Hence if there is unpronounced material here (at least material that is in any way relevant to the interpretation of the sentence) it must be *I invited (t)* and not *you invited (t)*. Example 37 makes the same point.

- (37) A: Who likes me?  
B: Bozo.

Here the unpronounced material is *likes you* (or, in a theory with traces and movement, *t likes you*) and not *(t) likes me*. This fact is obvious and has been known for decades, but it is ignored surprisingly often in the ellipsis literature.

Note crucially that it is not simply that there is a switch in understanding from first person to second person and vice versa. Consider the scenario in 38.

---

matching. Regardless of one’s theory of how the answer comes to ‘mean’ what it does, there would be a semantic mismatch (the object denoted by an accusative NP cannot be ‘helped’). But it is very widely assumed that, although there is some core of case-marking predictability both within and across languages, in the end there are idiosyncrasies that preclude a fully semantic account.

- (38) Speaker A to Speaker B (a female): ‘What state was your mother born in?’  
 Speaker B looks puzzled, unable to remember.  
 Speaker C is an onlooker and is not related to B, but jumps in and says: ‘Oh, I know. Illinois.’

Here C’s answer is understood as the proposition that B’s mother was born in Illinois, not that either A’s or C’s mother was. So the silent material here—if there is any—would have to be *Her mother was born in t* (or even *B’s mother was born in t*). But formal identity would require the silent material to be *Your mother was born in t*. As one more possibility, take 39.

- (39) [Context: Bristol Palin is talking to a reporter, and says:]  
 Who did my mother endorse yesterday?  
 [You and I overhear this. I turn to you (not to Bristol) and say:]  
 Actually, I know—Newt.

Here the proposition understood by the answer is that Bristol’s mother endorsed Newt, and so the silent material would have to be *her mother endorsed t* or *Bristol’s mother endorsed t* (or some such thing). But the silent material required by formal identity would be *my mother endorsed t*, which is obviously incorrect.

The point has been belabored in order to ward off some possible attempts to save the formal identity requirement that various people have suggested to me. (i) Could this be just a special instance of Fiengo and May’s ‘vehicle change’? The answer is no, for that proposal requires a simple featural change of –pronominal to +pronominal. Here we find a second-person feature changing to a first-person feature in 35, a first-person feature changing to second-person feature in 37, a second-person feature changing to a first-person feature in 38, a first-person feature changing to a third-person feature in 39, and the interested reader can construct the other logical possibilities. There is no single recipe that one could give for a feature change without consulting the value of the indexical in the given speech context. (ii) A referee suggests that perhaps this is like the case of allowing antecedent material and ellipsis sites to differ in terms of the indices on pronouns that they contain. (This of course presupposes that there are indices on pronouns, an assumption not being endorsed here. Nonetheless, I consider this possibility to see if, in a theory with indices, the indexical mismatches can be accounted for.) But this will not do. In the first place, it is not clear that indices on pronouns (in those theories that make use of these devices) are at all like person features. But even if they are, we are not just stripping off something: there has to be a systematic way to correspond the value of *I* to the value of *you*, the value of *you* to the value of *I*, and so forth. (iii) Along these lines, another referee suggests that this is just akin to the case of ‘sloppy identity’ whereby the elided material in 40 can be understood as John voting for John’s mother.

- (40) Bill<sub>i</sub> voted for his<sub>i</sub> mother. John did too.

But as has been known since Keenan 1971, so-called ‘sloppy identity’ itself can be seen as just identity of meaning. Since *his* in the first clause can be a bound pronoun, the meaning of the VP (or of some LF constituent under some theories) is  $\lambda x[x \text{ vote for } x\text{'s mother}]$ , and that is the understood meaning in the second clause. Just how this works, of course, itself depends on one’s account of VPE, but both SLM accounts and non-SLM accounts such as that in Jacobson 2003 have no problem getting the sloppy reading from independent mechanisms for pronoun binding. It should be noted that whether sloppy identity is handled by an identity-of-meaning or an identity-of-form condition is itself theory-dependent. Sag 1976 took the relevant condition to be identity of form at LF, but required formal identity up to ‘alphabetic variance’ and allowed the

antecedent clause to have a formula like  $\lambda x[x \text{ vote for } x\text{'s mother}]$  as part of its LF, while the elided VP could have the LF  $\lambda y[y \text{ vote for } y\text{'s mother}]$ . But even if one assumes that ‘sloppy identity’ is handled by a formal identity condition allowing a difference in variable names (under alphabetic variance), it is difficult to see how this has anything to do with the systematic shift in the interpretation of the indexicals in the examples above. Clearly, then, what is going on is that the understanding of the pronouns in the question and in the answer is such that they involve identity of meaning—more specifically, identity of CONTENT in the Kaplan 1989 terminology discussed earlier.

And as also discussed in Hankamer & Sag 1984, it is not just person features on indexicals that shift in the systematic way shown above. Suppose A is in Vermont and is talking on the phone to B who is in Paris where B has access to a wonderful weather forecast program.

(41) A: When is the snow supposed to start here tomorrow?

B: At three in the afternoon.

Obviously, *here* is not a silent part of B’s answer. Similar facts hold with temporal indexicals.

(42) [Context: A sends an email to B on Tuesday inquiring about the party B is planning for Wednesday. The email reads: ‘Who all will be coming to your party tomorrow?’ B doesn’t read the email until the next morning, but can certainly answer with:]

All of the students in my class and the three deans I want to impress.

Clearly, then, the requirement imposed by Shh must be that its complement have a MEANING—not a form—that is identical to the meaning of some other linguistically overt expression in the discourse context (in this case, part of the question).

But this is actually oversimplified. It is not just ‘meaning’ in the sense of a function from speech contexts to other sorts of model-theoretic objects—that is, it is not the ‘character’ in Kaplanian terms (recall the discussion of character vs. content in §2). Rather, it is the ‘content’, that is, a model-theoretic object that results after the values of the indexicals are fixed. To formalize what is going on here, let us use  $\llbracket[\alpha]\rrbracket$  to denote the character of  $\alpha$ —that is, the function from speech contexts to something else. Thus basing the discussion heavily on that of Hankamer and Sag (1984) but using Merchant’s condition on the sister of Shh, we can state this as follows.

(43) Let  $\alpha$  be the sister of Shh. Then there must be a context C with two subcontexts  $C'$  and  $C''$  such that there is a linguistic expression  $\beta$  in C such that  $\llbracket[\alpha]\rrbracket(C) = \llbracket[\beta]\rrbracket(C'')$ .

The idea here is that in a given discourse context C we can have two subcontexts  $C'$  and  $C''$ , which differ in terms of who is the speaker and who is the hearer (a natural consequence of normal dialogue). Then an expression  $\alpha$  can be the sister of Shh if the character of  $\alpha$  is such that its content in  $C'$  (where one person is the speaker) is the same as the content of the ‘antecedent expression’ in  $C''$  (where another person is the speaker). Thus it clearly is identity of meaning—in the sense of ‘content’—that is relevant for Shh.

In sum, even though there are some proposals in the literature to allow slight differences in form between antecedent and ellipsis (e.g. ‘vehicle change’), it is difficult to see any way one can allow the formal mismatch with respect to indexicals without consulting their meaning. Indexicals seem to be announcing quite forcefully that formal identity is not relevant; what is relevant is identity of meaning—of ‘content’.

Before returning to case matching, we should assure ourselves that the data concerning indexical shifts present no difficulty for the Qu-Ans analysis. Indeed they do not;

the semantics of Qu-Ans need only ensure that it is the CONTENT and not the character of the question that combines with the content of the answer.

RETURNING TO CASE MATCHING. But it turns out that giving up formal identity means giving up SLM as an explanation for case-matching effects. The point is easiest to initially illustrate in the abstract. Thus, imagine a language just like German but with a mythical verb *felfen* that is synonymous with *helfen* but that happens to assign accusative case. Then SLM makes no prediction about case matching, since the material that is missing could contain the verb *felfen* rather than *helfen* and the semantic match would still be ensured. And indeed, cases like this mythical one do exist. So, for example, the Hungarian verb *hasonlít* ‘resemble’ optionally assigns either sublativ (SUBL) case or allative (ALL) case to its object, as shown in 44.<sup>19</sup>

- (44) a. Péter hasonlít János-hoz.  
Peter resembles John-SUBL  
‘Peter resembles John.’  
b. Péter hasonlít János-ra.  
Peter resembles John-ALL  
‘Peter resembles John.’

Three informants have confirmed that there is no meaning difference between these two variants (we return to that below). But the short answer must nonetheless match the WH-word in case.<sup>20</sup>

- (45) a. Q: Ki-re hasonlít Péter?  
who-SUBL resembles Peter  
‘Who does Peter resemble?’  
b. A: János-ra.  
John-SUBL  
c. A: \*János-hoz.  
John-ALL  
(46) a. Q: Ki-hez hasonlít Péter?  
who-ALL resembles Peter  
‘Who does Peter resemble?’  
b. A: \*János-ra.  
John-SUBL  
c. A: János-hoz  
John-ALL

<sup>19</sup> I am extremely grateful to Balasz Suranyi for supplying me with this case and with the judgments of five speakers (see n. 20); I checked these with two additional informants. I had originally (Jacobson 2009) constructed this case with the Hungarian verb *segít* meaning ‘help’, which assigns both dative and accusative case, and which was suggested to me by Zoltán Szabó. Sag and Nykiel (2011) extended my argument to sluicing also basing their data on the verb *segít*. However, Suranyi pointed out to me that the argument based on *segít* has additional complications in that there are some slight meaning differences (centering on telicity) between the two. One can nonetheless construct minimal pairs (differing only in the case marking) in which the two are truth-conditionally the same, but the use of *hasonlít* provides a much cleaner test case and allows for the relevant construction using much simpler sentences.

<sup>20</sup> There is a slight complication. Balasz Suranyi asked five speakers (including himself) about the judgments in 45–46. All five agreed that the mismatch short answer in 46b is completely impossible. Three found the mismatch answer in 45c to also be completely impossible, but two (including himself) found this mismatch to be degraded but better. Two additional informants that I consulted found all of the mismatch cases to be bad. Why some speakers find improvement in 45c is unclear.

Note that matching is not required in the long reply; either of the full sentences in 44a and 44b can be used as replies to either of the questions in 45a and 46a (although the mismatch one might seem awkward or slightly degraded).

In fact, similar cases can be constructed for English using pairs like *promise* and *make a promise (to)*. While there may be subtle differences in the appropriateness conditions for these expressions, the two seem to make equivalent contributions to the truth conditions. Yet notice the impossibility of the question/short answer pair in 47a–b, even though the long replies in 47c and 47d—while decidedly odd in the discourse—are possible.<sup>21</sup>

- (47) a. Q: Who did Mitt promise that he would repeal the Affordable Care Act?  
 b. A: \*To the assembled tea party crowd.  
 c. A: Mitt made a promise to the assembled tea party crowd that he would repeal the Affordable Care Act.  
 d. A: To the assembled tea party crowd, Mitt made a promise that he would repeal the Affordable Care Act.

One can undoubtedly find many similar pairs in English with light verb + nominalization vs. full verb.

Let us take stock of the implications of these facts. First, as discussed at the outset of this section, the fact that the case and/or category of the short answer must match the case/category of the WH is immediate in the Qu-Ans account. For here the syntax of Qu-Ans directly builds this in: something is a Qu-Ans pair only if this matching obtains. And it is only if something is genuinely a Qu-Ans that the compositional semantics specifies that the meaning of the question and the meaning of the answer are put together in the way shown earlier to result in the propositional information. The point is that Qu-Ans is a linguistic unit specified by the grammar with both a syntax and a semantics—and case/category mismatch violates the syntax. Notice that the ‘identity condition’ for the syntax is on the OVERT material (the WH-word and its answer). In contrast, on the SLM approach the identity condition holds between what is silenced and its antecedent—not on what is overt. Moreover, there is no question here that the identity of the category/case of the WH-word and the answer is formal (syntactic) identity because this matching requirement is part of the syntax of the construction.

But what about SLM? The case-matching requirement on the answer and on the WH-word would follow (indirectly) under SLM only if the basic matching requirement (i.e. the conditions for Shh) require formal identity. The mismatch in 45a (sublative in question) and 45c (allative in short answer) would then follow, for the relevant portion of the

<sup>21</sup> Of course, Merchant’s account would not have 47c be the appropriate ‘source’ for 47b in any case. Rather, the ‘source’ is a sentence with fronting. Incidentally, the topicalized sentence in 47d is indeed quite odd in the discourse, but this has nothing to do with the fact that there is a mismatch between the relevant portion of the question (*Mitt promised* [<sub>NP</sub> *t*] *that he would repeal the ACA*) and the sister to Shh in the answer (*Mitt made a promise* [<sub>PP</sub> *t*] *that he would repeal the ACA*). The case with full matching is also odd.

(i) ?The assembled tea party crowd, Mitt promised that he would repeal the ACA.

The oddness presumably derives from the discourse conditions on topicalized constructions, which are very different from the conditions on replies to questions. Incidentally, one might think that the oddness of the full answer with fronting poses a problem for Merchant’s account (where material is obligatorily fronted in the answer) but would be mistaken. In Merchant’s account, the answer involves fronting material to the Spec position of a node labeled FP; its head is what I have renamed Shh, and Shh requires its complement to be silent. Overt topicalization cases like (i) presumably have the fronted material in Spec position of some other category—say, TopP—and the interpretation of TopP is different from that of FP and is appropriate only in certain discourse situations.

question would be [*Péter hasonlít t<sub>SUBL</sub>*], while the answer would be 48 (illustrating with a Merchant-style analysis).

(48) János-hoz [<sub>FP</sub> Shh [*Péter hasonlít t<sub>ALL</sub>*]]

But the sister of Shh has no formally matching antecedent, resulting in ill-formedness.

Yet we have seen that the identity cannot require formal identity (see also Merchant 2001 for this claim based on other kinds of evidence). We should, then, return to the question of whether there might be some slight difference in the meaning of the antecedent [*Péter hasonlít t<sub>SUBL</sub>*] vs. [*Péter hasonlít t<sub>ALL</sub>*]. For example, it could be that there are two closely related homophonous verbs here but with slightly different meanings, each of which assigns a different case.<sup>22</sup> But informants agree that there is no truth-conditional difference here; thus both of the following are contradictions.

- (49) a. #*Péter hasonlít Jánoshoz, de nem hasonlít rá /Jánosra.*  
 Peter resembles John.ALL but not resembles he.SUBL/John.SUBL  
 ‘Peter resembles John.ALL, but he doesn’t resemble him/John.SUBL.’  
 b. #*Péter hasonlít Jánosra, de nem hasonlít hozzá /Jánoshoz.*  
 Peter resembles John.SUBL but not resembles he.ALL/John.ALL  
 ‘Peter resembles John.SUBL, but he doesn’t resemble him/John.ALL.’

The same holds for the English case; 50a and 50b are contradictions.

- (50) a. #Mitt made a promise to the assembled tea party crowd that he would repeal the ACA, but he didn’t promise them that he would (repeal the ACA).  
 b. #Mitt promised the assembled tea party crowd that he would repeal the ACA, but he didn’t make a promise to them that he would.

Note that any semantic identity condition under SLM would have to be on the truth-conditional aspects of meaning;<sup>23</sup> it clearly needs to ignore extra bits of meaning such as implicatures about the speakers’ attitude. Thus such an implicature is stripped away in the short answer in the dialogue below.

- (51) Q: Who did that idiot Bozo invite to the party?  
 A: Claribel. But I don’t agree that Bozo is an idiot.

This shows that an identity condition would need to be formulated to care only about truth-conditional aspects of the meaning of the antecedent and the silenced part of the answer (although again see n. 23).

The bottom line, then, is that while the case-matching requirement has often been taken as one of the prime arguments for SLM, it is actually problematic for the SLM account. But it is straightforward under Qu-Ans. Of course, this is not to say that SLM could not have a strategy similar to the Qu-Ans strategy—it could be elaborated so as to also directly require matching between the WH-constituent in the question and that in the answer. But adoption of such a strategy undermines this particular argument FOR

<sup>22</sup> This appears to be a possibility with respect to the verb *segít* discussed in n. 19.

<sup>23</sup> This is oversimplified. AnderBois (2014) shows that sluicing, for example, is sensitive not only to truth conditions but also to what is ‘at issue’ and what is not. But there is no reason to think that the ‘at issue/not at issue’ divide would play any relevance in question-answer pairs (it is not clear how it would extend to this case). But even if it did, unless it could be shown that the difference between allative and sublative case, for example, caused there to be some at-issue difference, the fact that this additional dimension of meaning could play a role in the licensing of ellipsis would not help in explaining the case-matching facts. Similarly for the English case: it seems highly unlikely that the use of *promise* vs. *make a promise to* results in any difference in terms of the at-issue content.

SLM—and in fact it means that SLM would need two different identity conditions. The first requires identity between the antecedent and the silenced material, and the second requires identity between the remaining material and the WH-phrase—surely a rather complex set of conditions.

Finally, I should note that—under Qu-Ans—there is no tension between the semantic facts centering on the indexicals and the formal (category- and case-) matching requirement that holds between the WH-word and the answer. This tension is present under SLM, because here the indexical data are relevant to the formulation of a ‘matching requirement’ (since SLM posits that the short answer in 35b, for example, contains hidden material that must in some sense match a portion of the question). So the conundrum for SLM is that one set of facts appears to necessitate a formal identity condition while the other necessitates a meaning (content) identity condition. But under Qu-Ans the indexicals in the questions do not need to match anything else (indeed, there is nothing else for them to match!). Here, the only relevance of indexicals concerns the meaning (i.e. content) of the question; the semantics of Qu-Ans connects that with the meaning (content) of the answer. Of course, then, the particular form of an indexical is relevant only to the ultimate content (in a given context) that comes from that form.

**5. REPLYING TO (SOME) ARGUMENTS FOR SLM.** The above section presented three phenomena (the presupposition of the WH-word, answers to exhaustive questions, and case matching) that provide evidence for Qu-Ans over SLM. But the literature abounds with arguments for SLM, and so these must be addressed before victory can be claimed for Qu-Ans. Unfortunately, space precludes dealing with each and every one. The strategy, then, is to address a representative sample—hopefully representative and thorough enough as to make a reasonable case that the phenomena in question are all handled equally well without SLM. I address three broad classes of arguments here. The first (§5.1) centers on (apparent) syntactic evidence for any kind of SLM view. The second (§5.2) concerns arguments specifically for the movement-plus-ellipsis analysis of Merchant (2004). These have been used to indirectly argue for SLM because there cannot be movement unless there is SLM from which something has moved. Only one argument of this type is discussed in the main article, but two others—based on preposition stranding and island effects—are responded to in the online appendices. The third apparent argument for SLM (§5.3) is often implicit rather than explicit. This centers on a kind of ‘commonsense’ view of how it is that short answers (and elliptical constructions in general) are understood in the way they are. For each of these broad classes of arguments, I show that they rely on questionable assumptions and/or that—on closer inspection—the logic of these arguments does not go through.

**5.1. CONNECTIVITY ARGUMENTS.** Quite a number of the arguments for the SLM approach—going back to at least Morgan 1973—are of the following form (many of these have to do with ‘binding’ and/or so-called BINDING THEORY effects).

- A. PREMISE: Phenomenon P has to be stated with respect to some property of (nonlocal) chunks of representation.
- B. FACT: We find phenomenon P in fragment answers.
- C. CONCLUSION: Therefore there must be a level of representation at which the fragment answer is surrounded by other material.

The fallacy here lies in the premise, that is, in the assumption that we need access to global chunks of representation to account for the relevant phenomenon. Such phenomena include so-called ‘principle A’ effects, the related c-command constraint on the

binding of ‘picture-noun reflexives’ (which may or may not fall under principle A), principle B effects, principle C effects, and the standard mantra that ‘binding requires c-command’. But, as noted earlier, under DC these kinds of constraints on representations could in any case NOT be stated, and so the premise in A would have to be incorrect. Of course, this is of little consolation to an advocate of DC if there really is good evidence for the premise in A—that is, if it is difficult or impossible to account for the relevant phenomena in some other way. But indeed, in almost all of the relevant cases there are simple DC alternatives—alternatives that do at least as well and in some cases arguably better than the representational constraint approach.<sup>24</sup> And in fact, if there is evidence against the SLM view of fragment answers (i.e. if C is incorrect), then this, combined with B, shows that the initial premise in A must be incorrect.

I pick a representative sample of connectivity arguments: those based on so-called principle A effects, on bound-variable connectivity, on the binding of picture-noun reflexives, and on the distribution of subject reflexives. I do not discuss principle B effects, but the interested reader can verify that the approach to such effects developed in Jacobson 2007 accounts for the relevant effects in ellipsis.

REFLEXIVES. I begin with the distribution of reflexives. The argument for SLM is based on the assumption that 52b is good for the same reason that 52c is, and that 53b is bad for the same reason that 53c is, and that—crucially—the goodness of 52c and badness of 53c (in this context) CAN ONLY BE STATED WITH REFERENCE TO A FULL SENTENCE. (Here and throughout I use indices merely as a way to indicate relevant readings without any commitment to the view that the grammar makes use of indices).

- (52) a. Q: What did Mitt<sub>i</sub> do all night during the debate?  
 b. A: Praise himself<sub>i</sub>.  
 c. A: Mitt<sub>i</sub> praised himself<sub>i</sub>.
- (53) a. Q: What did Mitt<sub>i</sub> hope that Anne would do (on her campaign tour)?  
 b. A: \*Praise himself<sub>i</sub>.  
 c. A: \*Mitt<sub>i</sub> hoped that Anne would praise himself<sub>i</sub> (on her campaign tour).

The argument that the answer must covertly be the longer sentence stems from a very entrenched view of reflexives going back (although in a slightly different form) to at least Postal 1964 and slightly recast in the so-called binding theory (Chomsky 1981) under the rubric of principle A. The binding theory account holds that reflexives (and other NPs) come with an index, and that the grammar contains some constraint to the effect that a reflexive (at least one in object position) must be coindexed with a locally c-commanding NP.<sup>25</sup>

But a constraint like this is in any case incompatible with DC, and in fact there also is no need to make the principle A assumption. There are accounts of reflexives—such as Keenan 1971, Partee & Bach 1981, Szabolcsi 1987, and Steedman 1996—that differ in

<sup>24</sup> This article answers the arguments based in principle A and the ‘c-command condition on binding’, but space precludes any discussion of principle C effects. Suffice it to say that there are accounts of so-called principle C effects that locate the effect in discourse principles (see e.g. Kuno 1975), and these require no reference to a coindexation constraint stated on some chunk of representation.

<sup>25</sup> The account given in Postal 1964 and assumed in some form in much of the early work in transformational grammar did not contain a constraint on representations per se. Rather, there was an obligatory rule, changing an ordinary NP (or pronoun) into a reflexive NP if it was in the same local domain (i.e. same S) as a previous coreferential NP. Since reflexive pronouns were not present at deep structure, this is the only place they were allowed. But although this did not use a filter on representations—as has become common—it nonetheless required reference to nonlocal chunks of representation (e.g. the domain containing a subject and an object).

detail but that all have the end result that the meaning of a VP like *praise himself* is  $\lambda x[[\text{praise}]](x)(x)$  (i.e. the characteristic function of the set of self-praisers). From this, it follows that that 53c cannot have the reading indicated by the indices since (ignoring the auxiliary)  $[[\text{praise himself}]]$  ultimately takes the individual named by *Anne* as argument. And 53b will be bad (on the intended reading) for the same reason.  $[[\text{praise himself}]]$  is an  $\langle e,t \rangle$  function denoting the set of self-praisers and, in the Qu-Ans analysis, this combines with the meaning of the question that is the function represented informally in 54.

(54)  $\lambda P_{\langle e,t \rangle}[\text{Mitt hoped that } P(\text{Anne})]$

In the end, then, the Qu-Ans pair conveys that Mitt hopes for the proposition that Anne has the self-praising property: a perfectly good meaning (modulo the gender conflict if one assumes that Anne is female), but not the reading indicated by the indices in 53b. (Note that I assume here that gender is semantic, and so  $[[\text{praise himself}]]$  is actually the function characterizing the set of self-praisers but restricted to the domain of males. If it is syntactic, there are also ways to handle the gender mismatch via feature passing.)

INTERLUDE: CONNECTIVITY IN SPECIFICATIONAL SENTENCES. It is worth briefly pausing to note that the arguments like that given above (centering on principle A) for the SLM analysis of short answers have also been treated in detail in the related literature on connectivity effects in specificational sentences. Thus, as discussed extensively since at least as early as Higgins 1973, we find the same sorts of ‘connectivity effects’ in specificational sentences, including both those whose precopular constituent is a free relative (i.e. those traditionally referred to as ‘pseudo-clefts’) and those whose precopular constituent is a fully headed NP.

- (55) a. What Mitt<sub>i</sub> did (during the debate) was praise himself<sub>i</sub>.  
 b. \*What Mitt<sub>i</sub> thought Anne would do (on the campaign trail) is praise himself<sub>i</sub>.
- (56) a. The most obnoxious thing that Mitt<sub>i</sub> did (during the debate) was praise himself<sub>i</sub>.  
 b. \*The most obnoxious thing that Mitt<sub>i</sub> thought Anne would do is praise himself<sub>i</sub>.

The reflexive facts have at times been taken to necessitate an SLM analysis of the postcopular constituent, according to which it is actually *Mitt praised himself* (or *himself* [*+Mitt praised*]). Ross (1985) and Schlenker (2003) explicitly liken specificational sentences to question/answer pairs, where the precopular constituent is a concealed question and the postcopular constituent is its answer. Further, they simply assume the SLM analysis of short answers, and so the existence of silent material in the postcopular constituent in specificational sentences merely reduces to the existence of such material in short answers. Hence we have simply come full circle. While it may well be that specificational sentences do bear a relationship to question-answer pairs, the question of whether the postcopular constituent in the former contains silent material reduces (possibly) to the question of how to analyze short answers in general. I therefore do not consider specificational sentences any further, except to note that there are already a number of papers showing that the apparent connectivity effects in specificational sentences do not require SLM, and some of these alternative analyses will form the basis of the subsequent discussion here of other connectivity effects (see e.g. Jacobson 1994, Sharvit 1999, and Heller 2002).

BOUND-VARIABLE CONNECTIVITY. There is a common adage in work on binding, which goes as follows: ‘A pronoun must be c-commanded (at LF) by its binder’. I am

calling this an ‘adage’ because often this is taken to be not a constraint in the grammar along the lines of, for example, principle A, but rather an observation about how the semantics works. The assumption is that 57 (on the reading indicated by the indices) is derived via an LF in which *every third grade boy* is ‘raised’ to give the LF in 58 (this particular way of notating the LF is essentially the notation in Heim & Kratzer 1998).

(57) Every third grade boy<sub>i</sub> called his<sub>i</sub> mother.

(58) [every third grade boy] [<sub>Λ</sub> 8 [<sub>S</sub> t<sub>8</sub> called his<sub>8</sub> mother]]

Continuing to exposit the common adage, this is based on a view of grammar that makes use of indices and variables: the idea is that every expression has a semantic value relative to a way to assign values to the variables. Call each such way an assignment *g*; hence any *g* is a function from variable names to individuals. Then  $\llbracket \text{his}_8 \text{ mother} \rrbracket$  on any *g* is the individual who is the mother of *g*(8);  $\llbracket \text{call his}_8 \text{ mother} \rrbracket$  on *g* is the set of individuals who called the mother of *g*(8); and  $\llbracket t_8 \text{ called his}_8 \text{ mother} \rrbracket$  on any *g* is the proposition that *g*(8) called the mother of *g*(8). The technique for getting the ‘bound’ reading is built into the next step (call this the ‘binding’ step). Thus the value of the node labeled  $\Lambda$  in 58 is a constant function from the set of assignments to a function of type  $\langle e, t \rangle$ —each *g* is assigned the function  $\lambda x [x \text{ called } x\text{'s mother}]$ —that is, the set of self’s-mother-callers. (Were there another pronoun within the lowest *S* that was still unbound, the value of the LF expression headed by 8 would not be constant on all assignments, but would still vary.) Very briefly, the way that this value is computed is that an expression like [<sub>Λ</sub> 8 *S*] on any *g* is a function from individuals *x* to the value of *S* on the assignment just like *g* except where 8 is assigned to *x*. This then combines with the meaning of the raised generalized quantifier; on any *g* the value of [8 [<sub>t<sub>8</sub></sub> called his<sub>8</sub> mother]] is taken as argument of  $\llbracket \text{every third grade boy} \rrbracket$  on *g*.

This is indeed one possible theory of the compositional semantics and how it gives rise to the ‘bound’ reading. It makes use of indices in the syntax that correspond to variables in the semantics, it makes use of the notion of assignments (functions from the integers to individuals), and it makes use of a ‘binding’ step that is the step that gives the meaning of the expression labeled  $\Lambda$  in the above LF. And if indeed this were the only way to account for ‘bound’ readings, it would appear that the answer in 59b must contain additional material (which includes the subject *every third grade boy*, which would undergo QR) so as to be mapped to the LF in 58.

(59) a. Q: Who did every third grade boy<sub>i</sub> call (on his first day of school)?

b. A: His<sub>i</sub> mother.

But the assumption that this is the only way to set up a semantics to account for ‘bound’ readings is simply incorrect. There are alternative proposals regarding binding that are at least as simple (arguably, simpler). I sketch just one here, the one developed in Jacobson 1994, 1999 (and other papers), which uses a variable-free semantics. This makes no use of assignments as part of the semantic machinery, no use of indices in the syntax, and is fully compatible with a DC architecture. (Note that this semantics was not designed for the case of short answers to questions, but it happens to extend effortlessly to this; considerable independent motivation for this view can be found in Jacobson 1994, 1999, 2000, among others.)

First, a pronoun such as *he* denotes the identity function on individuals—it is thus not of type *e* but of type  $\langle e, e \rangle$ . Now consider the composition of sentences like 60.

(60) a. Every boy<sub>i</sub> thinks that he<sub>i</sub> lost.

b. Every third grade boy<sub>i</sub> called his<sub>i</sub> mother.

Looking at 60a, the meaning of *lost* in the lexicon is of type  $\langle e, t \rangle$  and hence not the right type to directly combine with  $\llbracket \text{he} \rrbracket$ . But there is a simple mapping from any function *f* of

type  $\langle a, b \rangle$  to a function of type  $\langle \langle c, a \rangle, \langle c, b \rangle \rangle$ , which is known in the categorial grammar literature as the ‘Geach rule’ or ‘Division’; I notate the relevant operation as  $\mathbf{g}$ . (This has unfortunate homonymy with the earlier (and standard) use of  $g$  as an assignment function. Here there are no assignment functions, and to make sure that the two uses of ‘ $g$ ’ are kept distinct I use boldface  $\mathbf{g}$  to indicate the ‘Geach rule’.) In particular, for any function  $f$  of type  $\langle a, b \rangle$ ,  $\mathbf{g}(f) = \lambda X_{\langle c, a \rangle} [\lambda C_c [f(X(C))]]$ . Informally, the input function  $f$  wants something in a set  $a$  to give something in  $b$ .  $\mathbf{g}(f)$  takes an ‘incomplete’  $a$ —that is, a function from  $c$  to  $a$ , and the result inherits the incompleteness. Thus Jacobson 1994, 1999 posits that the grammar contains a unary rule that maps  $\llbracket \text{lost} \rrbracket$  of type  $\langle e, t \rangle$  to  $\mathbf{g}(\llbracket \text{lost} \rrbracket)$  of type  $\langle \langle e, e \rangle, \langle e, t \rangle \rangle$ .  $\mathbf{g}(\llbracket \text{lost} \rrbracket)$  takes a function  $f$  of type  $\langle e, e \rangle$  and an individual  $x$  and returns true if  $\llbracket \text{lost} \rrbracket$  maps  $f(x)$  to true. (In the system developed in the works cited above, the relevant unary rule also changes the syntactic category of the input and output item, but we ignore the syntactic details here.) When  $\mathbf{g}(\llbracket \text{lost} \rrbracket)$  takes  $\llbracket \text{he} \rrbracket$  (the identity function on individuals) as argument, the result is just the function  $\llbracket \text{lost} \rrbracket$ . (This is ignoring the contribution of gender; if  $\llbracket \text{he} \rrbracket$  is the identity function on male individuals, then  $\llbracket \text{he lost} \rrbracket$  is the lost-function defined only for males.)

Note then that any expression containing a pronoun that—informally speaking—is ‘unbound’ within that expression denotes a function from individuals to whatever is the type of meaning that corresponding ‘closed’ expressions have. Thus, *he lost* in 60a is a function of type  $\langle e, t \rangle$ . The full combinatorics for something like *his mother* in 60b require a digression into the semantics of genitives that would be too lengthy here; suffice it to say that in the ‘standard’ view with variables,  $\llbracket \text{his mother} \rrbracket$  is an assignment-dependent individual (and *his* must have some index), while here it ends up being a function from individuals to individuals—in particular, the function mapping each individual into that person’s mother.

The last key step is to provide an account of the ‘bound’ reading. Note that something extra is needed to get this in the account with variables as well. There, ‘binding’ is the result of the rule interpreting the node labeled  $\Lambda$  in 58, according to which the value of the embedded  $S$ —an assignment-dependent proposition—shifts to a constant function mapping each assignment to a function of type  $\langle e, t \rangle$ . In the variable-free system in Jacobson 1999, ‘binding’ is accomplished by a more local shift rule, which maps any function of type  $\langle a, \langle e, b \rangle \rangle$  to one of type  $\langle \langle e, a \rangle, \langle e, b \rangle \rangle$  and, informally speaking, ‘merges’ the two  $e$ -positions together. This rule is called  $\mathbf{z}$ . Thus for any  $f$  of type  $\langle a, \langle e, t \rangle \rangle$ ,  $\mathbf{z}(f) = \lambda X_{\langle e, a \rangle} [\lambda x_E [f(X(x)(x))]]$ . To illustrate,  $\llbracket \text{call} \rrbracket$  is the ordinary two-place relation (a function of type  $\langle e, \langle e, t \rangle \rangle$ ) holding between two individuals.  $\mathbf{z}(\llbracket \text{call} \rrbracket)$  is thus a function of type  $\langle \langle e, e \rangle, \langle e, t \rangle \rangle$  such that to  $\mathbf{z}$ -call some function  $f$  (of type  $\langle e, e \rangle$ ) is to be an  $x$  who ordinary-calls  $f(x)$ . This then completes the analysis of 60b;  $\llbracket \text{call} \rrbracket$  maps to  $\mathbf{z}(\llbracket \text{call} \rrbracket)$  and so wants as object a function of type  $\langle e, e \rangle$ . (Again this is also accompanied by a syntax that I do not develop here.) Since  $\llbracket \text{his mother} \rrbracket$  is just such a function, the meaning of *call his mother* is the function  $\lambda x [x \text{ called } x\text{'s mother}]$ —that is, the function characterizing the set of self’s-mother-callers—and this combines with the subject (in this case, it is taken as argument of the generalized quantifier *every third grade boy*). Similar remarks hold for the ‘binding’ step in 60a. Here,  $\llbracket \text{thought} \rrbracket$  undergoes  $\mathbf{z}$  to give  $\mathbf{z}(\llbracket \text{thought} \rrbracket)$ , which is a function of type  $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$  and is such that to  $\mathbf{z}$ -think some property  $P$  is to be an  $x$  who thinks that  $P$  holds of  $x$ . Thus,  $\llbracket \text{thought he lost} \rrbracket$  is the function characterizing the set of  $x$ ’s who thought that  $x$  lost.

It is worth pointing out again that this account of binding was not designed to account specifically for the case of short answers, and Jacobson 1999, 2000, and so forth provide independent motivation for this view. But strikingly, it immediately extends under Qu-Ans to question-answer pairs like 59 with no additional assumptions or machinery.

(The parallel case of specificational sentences is discussed in Jacobson 1994, and the discussion here basically recapitulates the points there.) Note first that, under most analyses, 59a is not an ordinary question but a functional question, discussed originally in Groenendijk & Stokhof 1984 and in Engdahl 1986. Both of these analyses take the meaning of 59a not to be a question about an individual, but rather a question about a function of type  $\langle e, e \rangle$ . Informally, it can be paraphrased as *what is the function  $f$  (of type  $\langle e, e \rangle$ ) such that every third grade boy is an  $x$  who called  $f(x)$* . In the accounts above this is accomplished by, roughly speaking, having the trace or gap following *call* translate as a ‘complex variable’—a variable  $f$  over functions of type  $\langle e, e \rangle$  applied to an individual variable.

In the variable-free framework under consideration here, the final semantics is the same. Strikingly, though, nothing extra is needed in order to get this semantics. Put differently, the existence of functional readings for questions is automatic and requires no extra apparatus (such as the complex traces used in some accounts). Consider an ordinary question like the reading for 59a in which the expected answer is an individual such as *the queen*. Assume (roughly) the account of extraction in Steedman 1987 in which *every third grade boy call* contains no trace, but is put together by function-composing *every third grade boy* with *call*. Leaving aside the full details of the meaning of the WH-word, assume *who (did) every third grade boy call* has essentially the same meaning as *every third grade boy call*. (As discussed earlier, the meaning of the WH-word will constitute a restriction on the domain of the function denoted by the question; for the purposes of the present discussion, it is not crucial to spell out just how this happens.) The interesting point here is the fact that the functional reading is an automatic consequence of the system: the functional reading just involves composing  $z(\text{call})$  with *every third grade boy*, and this gives as meaning  $\lambda f[\text{every third grade boy } z\text{-call } f]$ . Again, we assume that combining this with *who* yields the same meaning.

Not only does the existence of the functional readings for questions fall out automatically under the variable-free view, but the meaning of the fragment answer does also. Under the usual view, ‘binding’ requires *c-command*; the answer in 59b to a functional question would have to be a full sentence roughly like 60b—hence the argument for SLM. But notice that in the variable-free view nothing like this is required. The expression *his mother* automatically denotes a function of type  $\langle e, e \rangle$ . It is thus of the right type (and category) to combine with the meaning of the question to give a Qu-Ans, and the semantics of the Qu-Ans comes out just right.

To summarize: the argument for SLM based on ‘bound-variable connectivity’ is rooted in a particular view of how binding takes place, which may not be correct. Indeed, if DC is correct, this view of binding cannot be. But it turns out that there is a straightforward alternative account that not only allows the short answer to contain no silent linguistic material, but also has the additional advantage that the functional reading of questions comes ‘for free’. And, as noted above, considerable independent evidence for this analysis is detailed in Jacobson 1999, 2000, for example.

ADDITIONAL REFLEXIVES. It was noted above that reflexives in argument position in a VP (or an AP) can be treated in such a way that a VP like *praise himself* denotes the set of self-praisers. In that case, no SLM is needed to account for the dialogue in 52a–b. But this says nothing about other kinds of reflexives, such as the so-called picture-noun reflexives as in 61.

(61) John liked the picture of himself.

It is often assumed that these must be c-commanded (perhaps locally) by some coindexed NP. And if so, then the following question-answer pair must contain hidden material in the answer in order to satisfy this requirement.

(62) Q: What did John try to hide?

A: The picture of himself that had appeared in the newspaper.

But by now, considerable evidence that there simply is no such constraint on these reflexives has been amassed. Rather, as first discussed in Kuno 1975, there is some 'point of view' constraint concerning their role in a discourse, rather than a constraint stateable on syntactic structures. For extensive discussion with relevant examples, see Zribi-Hertz 1989 and Pollard & Sag 1992. The dialogue in 62 thus provides no evidence for SLM.

Similar remarks hold for bare reflexive answers as in 63.

(63) Q: Who did John vote for?

A: Himself.

For relevant discussion of the distribution of these in clefts see Pollard & Sag 1992; their remarks will carry over directly to the Qu-Ans case. Finally, Merchant (2004) notes that one must still account for the fact that bare reflexives are at best marginal as answers to questions where the WH-word is a subject. His particular examples are more complex than need be as they relate also to principle C effects (which, due to space limitations, are not discussed here). The basic pattern, however, can be shown by 64.

(64) Q: Who will punish him if he fails?

A: ?\*Himself.

A: \*Himself will punish him if he fails.

But this can be explained equally well via a case mismatch. *Who* is presumably NP[NOM], and it is reasonable to assume that *himself* (and other reflexive pronouns) inherently has an ACC feature on it.<sup>26</sup> Incidentally, it is well known that case mismatch is allowed (and required) for bare pronouns; thus only 65b and not 65c is a possible answer to 65a.

(65) a. Q: Who/\*whom won?

b. A: Him.

c. A: \*He.

<sup>26</sup> A referee questions this account by correctly noting that the same constraint holds for subject reciprocals, even though here there is no good reason to think these are necessarily case-marked [ACC]. Thus, just as (ib) is bad so is (ic), which might give rise to the suspicion that there really is a full connectivity effect going on here, necessitating SLM.

(i) a. Q: Which candidates voted for Barack and Hillary?

b. A: \*Each other voted for Barack and Hillary.

c. A: \*Each other.

But a closer look reveals that (ic) is bad for independent reasons. The corresponding case (where the 'gap' is in object position) is possible only if the question is understood as a functional question (or, perhaps, a pair-list reading).

(ii) a. Q: Which candidates did Barack and Hillary vote for?

b. A: Barack and Hillary voted for each other.

c. A: Each other.

But it is well known that functional (and/or pair-list) readings are impossible in any case when the gap is in subject position. See May 1985 and Engdahl 1988 for discussion.

By itself this fact is unaccounted for in both the SLM and the Qu-Ans accounts of case matching. As discussed in some detail in Ginzburg & Sag 2000, there appears to be a second principle overriding the case-match apparatus requiring the strong form *him* and not *he* in a variety of environments in English, and in particular in any environments where these stand alone. (One cannot, for example, point to someone and say: *Look! He!* Rather, it must be: *Look! Him!*) But there is no reason to believe that this constraint requiring the use of a strong form would be relevant for the reflexive matching requirement, and thus no reason to believe that the impossibility of 64 is due to anything other than a violation of the matching requirement.

**5.2. ARGUMENTS BASED ON THE PRESENCE OF MOVEMENT.** A second group of arguments for SLM (discussed especially in Merchant 2004 but found elsewhere as well) center on the claim that the fragment answer must be derived by movement followed by deletion/silencing. In other words, they argue for the SLM account sketched in §3.1. If indeed there is evidence that the short answer has moved from somewhere to a fronted position, then of course there must be surrounding material out of which it has moved.

**NONCONNECTIVITY CASES.** One such argument centers on cases where the short answer actually shows a category mismatch with the position of the ‘gap’ in the WH-question and/or the category of the WH-word. Merchant notes that these are exactly the mismatches found in topicalization or other constructions involving a dislocation, and so these are unsurprising on the movement-plus-silencing SLM account.

To elaborate, let us step back and look at this from the point of view of both the Qu-Ans account and the SLM account. The Qu-Ans account faces a surprising instance of mismatch from rather simple cases that were already given in 21 and are repeated here.

- (21) a. What silly thing that Donald heard from Sarah did he actually believe?  
 b. That someone put two phony birth announcements in Hawaiian newspapers to conceal the birth of a baby in Kenya.

This question-answer pair is impeccable, which I did not comment on earlier, but is actually surprising under Qu-Ans. Assuming that NP and CP are distinct categories (an assumption to be revisited below), then there is a category mismatch here. The WH-expression is obviously an NP—as witnessed by its internal structure containing a head noun *thing* and a relative clause—but the answer is a CP. Note though that this case is unproblematic under SLM, because SLM enforces no connection between the category of the WH-word and that of the answer. The necessary connection there is between the category of the short answer and the category that would be allowed in the position of the gap (i.e. the trace) following *believe* in the fuller structure in 66.

- (66) [<sub>FP</sub> [<sub>CP</sub> That someone ... in Kenya] [Shh [~~he actually believed~~ [<sub>CP</sub> *t*]]]]

This particular case causes no problem, because *believe* allows both NP and CP complements. In 21a the trace in the WH-question is presumably an NP trace. But if one assumes that there is no formal matching required between the material *he actually believed t* in the question and in the answer, this mismatch is unproblematic. (It is worth noting that if formal matching were required, then this case would be problematic for SLM as well since the question contains an NP trace and the answer presumably contains a CP trace.)

However, there are more elaborate cases along these lines for which more needs to be said under either account. Thus consider the distribution of complements of prepositions first noted in Rosenbaum 1967. It appears that prepositions take NP but not CP complements (Rosenbaum’s actual way of stating the generalization was a bit different).

- (67) a. \*I am ashamed of that I copied a lab assignment in college.

- b. I am ashamed of the fact that I copied a lab assignment in college.

Yet the following dialogue is impeccable.

(68) a. Q: What horrible thing are you (the most) ashamed of?

b. A: That I copied a lab assignment in college.

Under Qu-Ans, the surprise is just the same as what we saw in 21: the categories of the fragment answer and of the WH-expression do not match. Under SLM (with semantic rather than formal identity), the surprise is not about matching *PER SE*. Rather, the ‘surprise’ revolves around the ‘silent’ portion of 68, which must be ~~*I am ashamed of [NP CP]*~~, where the trace must be NP and yet the moved material is CP.

But Merchant actually parlays this mismatch into an argument FOR SLM, because we know that indeed this kind of mismatch is allowed (or at least much improved) in other dislocated positions, as in topicalization.

(69) That I copied a lab assignment in college, I am (most) ashamed of.

Of course, it remains to be explained WHY such mismatches are allowed (we return to this below), but given that they are Merchant argues that the movement-plus-silencing account predicts mismatches of this sort.

Similar facts can be duplicated with a class of verbs discovered by Grimshaw (1982) that—for some speakers—surprisingly allow NP but not CP objects; a good example is *capture*. Here too (in the relevant dialect) NPs but not CPs are allowed in object position, but CPs are fine in both topicalized positions and short answers.

(70) a. No theory has yet captured the fact that there are black holes.

b. \*No theory has yet captured that there are black holes.

c. That there are black holes no theory has yet captured.

d. Q: What amazing fact has no theory yet captured?

A: That there are black holes.

Another such case (which holds for more speakers) centers on the verb *dislike*. This verb, as opposed to *like*, is not happy with an infinitive VP (or S, depending on one’s theory) complement. Again, though, the violation disappears in both topicalization and short answers.

(71) a. I dislike the thought of losing at poker.

b. I \*dislike/like to lose at poker.

c. To lose at poker, I really dislike.

d. Q: What do you dislike (the most)?

A: To lose at poker.

Once again, if the short answer is derived by movement followed by deletion, the existence of these category mismatches is unsurprising. The mismatch between the WH-expression and the answer is unproblematic because SLM enforces no matching requirement here; the mismatch between the position of the trace in the silenced portion and the category of the fronted material is unsurprising because the violation in general disappears under movement.

But the argument is flawed. The fact that two constructions B and A share certain properties with respect to some phenomenon P in no way shows that B COMES FROM A. That conclusion follows only if there is reason to believe that the relevant generalization could only be stated with respect to properties that A has and that B superficially does not have. (Note that in Merchant’s analysis an answer like 68b does not actually come from the sentence in 69, for the latter is not an FP. Not only are the discourse conditions different, but the head of FP is *Shh*, which necessarily suppresses the phonology of its sister. However, Merchant is not actually claiming that 68b comes from 69; his point is that the violation

disappears under ‘dislocation’ in general, and so as long as there is movement in an FP the facts follow.) Any theory thus needs an account of why the violation is found only in some places and not others, so there could well be a single explanation for the goodness of the short answer and the topicalized case without there being any derivational relation between these. Indeed, this is exactly the tack I take here.

Let us first reconsider just why it is usually assumed that NP and CP are different categories. For one thing (a fact already noted in Rosenbaum 1967, although using different language), CPs are allowed in environments where NPs are not. This includes the complement of a verb like *pray* and the complement of adjectives (and nouns).

- (72) a. I pray that the election is not a disaster.  
 b. \*I pray the hoped-for outcome.  
 c. \*I pray something.
- (73) a. I am really happy that he won.  
 b. \*I am really happy the fact/outcome/proposition that he won.  
 c. \*I am really happy something.

(The (c) sentences are included here to reassure ourselves that the problem is not semantic, since *something* is a very ‘bleached’ NP with little restrictions on what it can range over.) But Rosenbaum (despite his observation about prepositions) assumed that CPs could nonetheless occur wherever NPs can. That is, (modulo the situation with prepositions) if an environment allows for an NP it also allows for a CP, although not necessarily vice versa. So Rosenbaum accounted for this by proposing a phrase structure rule to the effect of NP → CP. (This was not his actual rule, but this is a reasonable, more modern reconstruction of his rule.) Or, to put this in terms of the DC architecture sketched in §2, this idea can be resurrected in the form of a unary rule to the effect that any expression  $\alpha$  of category CP maps to one with the same meaning and phonology but with category NP. The empirical prediction, then, is that anywhere an NP can occur so can a CP (provided that it is an environment compatible with a proposition-denoting expression) but not vice versa. This could be extended to the case of infinitive VPs; imagine that there is a rule allowing them to shift category to NP.

Now of course this immediately runs counter to the facts directly above: we see that there are environments allowing NPs but not CPs (nor infinitive VPs). These include not only objects of prepositions, but also the complement position (for some speakers) of *capture* (and similarly for *dislike*). For the moment, though, let us set this fact aside. If CPs really are also NPs, then there is no mismatch between the gap position and the dislocated constituent in topicalization: both are NPs. And there is no mismatch between the short answer and the WH-expression in 21, 70d, and 71d—in all cases we are dealing with NPs. Problem solved, and the solution holds equally well for Qu-Ans as for SLM.

But of course we cannot just stop here, for this blatantly ignores the problem with prepositional objects and the object of *capture* and *dislike*: these allow NPs but do NOT permit CPs (and infinitive VPs in the case of *dislike*). But suppose we reframe the problem not as ‘Why does the violation disappear where it does?’ but as ‘Why does the violation appear where it does?’. Then there is an obvious answer: the violation appears only under adjacency. Indeed, Rosenbaum noticed this with respect to the prepositions

<sup>27</sup> I thank Arnold Zwicky for first pointing this solution out to me regarding the *capture* class; Rosenbaum himself observed it for the preposition case.

<sup>28</sup> There are two concerns about how to state this direct-compositionally. The first is how exactly to build in the adjacency requirement itself (although the interested reader can imagine various possibilities). The sec-

and proposed an obligatory preposition-deletion rule if the preposition is adjacent to the CP. This cannot be the solution since it does not extend to the *capture* case, but the basic insight—that it has to do with adjacency rather than pure selection—seems to be absolutely correct.<sup>27</sup> To be sure, I have no account of just how to state this restriction in a DC framework, and so I leave a full account as a promissory note.<sup>28</sup> But there is indeed evidence that adjacency is what is relevant, for there are many ways to remove adjacency and remove the violation. As to the situation with CP complements of prepositions, consider the following striking contrasts.

- (74) a. \*I have always regretted and continue to be ashamed of that I copied a lab report in college.  
 b. I have always been ashamed of and continue to regret that I copied a lab report in college.

Under most theories of ‘right node raising’ constructions like this, the basic structures of 74a and 74b are the same; only the word order of the two conjoined parts differs. Similar facts regarding coordination can be constructed for *dislike* (the same pattern holds for *capture*; I leave it to the reader to verify).

- (75) a. \*I used to absolutely hate but now only mildly dislike to lose at poker.  
 b. I now only mildly dislike but I used to absolutely hate to lose at poker.

Or take the following, where an adverb intervenes.

- (76) a. His theory captures—with great elegance and simplicity—that the universe contains black holes.  
 b. I violently dislike—with every fiber of my being—to lose at poker.

Thus the mismatch in topicalization is accounted for as well: the offending material is not adjacent to the verb. And, given a fairly obvious definition of ‘adjacency’, there is no adjacency in the Qu-Ans case either. The two therefore do not need to be subsumed as the same thing; the reason that there is no violation is just the same in both cases. Of course, we do need to note that the definition of ‘adjacency’ has to mean ‘adjacency’ within a sentence and not across sentences in a discourse (or across speakers). After all, the expression *that I copied a lab assignment in college* is literally ‘adjacent’ to a preposition in the Qu-Ans pair in 68. However, it does not seem the least bit far-fetched to assume that the proper formulation of this condition in the grammar will not extend to cases of ‘adjacency’ across sentences/speakers. The upshot, then, is that Qu-Ans has no problem with the apparent ‘mismatch’ of CP and NP. The ‘CPs’ in question really are NPs, so there is category matching in the Qu-Ans. It is only when the relevant ‘NPs’ immediately follow a preposition or a verb like *capture* that there is a problem. And since any theory needs an account of the apparent mismatch in topicalization of the category of the ‘topicalized’ expression and the category asked for by *capture* (or a preposition), the suggestions here are not extra machinery cooked up just to save Qu-Ans.

---

and is that if we do accept the idea that expressions of category CP can map to NPs, then any attempt to state the requirement would appear to violate the locality restrictions inherent in DC. That is, an expression of category NP has no way to ‘know’ that it was licensed in virtue of the fact that an expression with the same phonology and meaning can be a CP. This problem, however, is easy to solve by simply assuming that these map to NP with some special feature (say NP [+CP]) and that the grammar will impose the adjacency requirement only on these kinds of NPs. Of course, one needs to also ensure that this very feature need not match on the WH-expression in a Qu and in an Ans in order to account for Qu-Ans pairs like 35. In general, it seems like not all features need to match, and a much more thorough investigation of categories and features is needed here to account for which ones really do need to match and which do not.

Before leaving this, I can briefly mention that a related argument given in Merchant 2004 (see also Merchant et al. 2013) regarding the presence or absence of the complementizer *that* can be answered in exactly the same way. Thus Merchant notes a kind of anticonnectivity effect in that *that* can be absent in embedded CPs when they are in situ, but not when they are topicalized or short answers to questions.

- (77) Ted believes (that) the earth is getting cooler.  
 (78) a. That the earth is getting cooler, Ted believes.  
       b. ?\*The earth is getting cooler, Ted believes.  
 (79) a. Q: What does Ted believe?  
       b. A: That the earth is getting cooler.  
       c. A: ?\*The earth is getting cooler.

(I have notated these with ?\* as they seem degraded but not impossible. Actually, 78b is good if read with the prosody associated with a parenthetical interpretation of *Ted believes*; presumably this has a different analysis from the topicalized version.) But there is again no reason to think that because both 78b and 79c are bad it must be the case that the latter comes from the former. They can be bad for the same reason, and that reason could be that the *that* can be absent only under adjacency with the verb. (For relevant discussion, see Bošković & Lasnik 2003, although the terminology and proposal are slightly different.)

ADDITIONAL ARGUMENTS BASED ON MOVEMENT. Merchant 2004 gives two other arguments for the analysis of movement followed by ellipsis. One concerns the so-called ‘P-stranding generalization’ according to which there is a correlation between whether a language has obligatory pied-piping and the possible form of the fragment answers—a correlation that supposedly follows under the movement analysis. The second concerns apparent island effects. Space precludes detailed discussion of these two arguments here, but see the online appendices. Here I just briefly outline the gist of those appendices. First, the facts are actually far from clear; it is well known that there are apparent counterexamples to the P-stranding generalization, and the ‘island’ facts appear not to actually line up with islands in general. Second, and perhaps more important, SLM actually does not account for either of these domains even if the putative generalizations were correct. As to the P-stranding cases, the full set of facts displays some surprising mismatches that present an interesting challenge to both SLM and to Qu-Ans (actually, the Qu-Ans analysis does slightly better, but neither theory accounts for the full range of facts). The so-called island effects do not follow under SLM because there is no obvious overt question to supply the identical material; the argument here is based on fragment answers to IMPLICIT QUESTIONS, not explicit ones. Just how to handle that is an interesting question for both theories. But since there is no actual spelled-out SLM account of these (nor is it clear how there could be), then even if the facts were as claimed they cannot provide evidence for SLM. And finally, SLM with a semantic identity condition additionally cannot account for the putative island generalization because nothing blocks the ‘bad’ cases from involving ellipsis of material containing a resumptive pronoun. All in all these domains are complex and not well understood, but—despite the fact that these are often cited as evidence for SLM—they do not provide any serious argument for this. Again I refer the reader to the details in the online appendices.

**5.3. THE PROPOSITION COMES FOR FREE. OR DOES IT?** A third apparent advantage of the SLM analysis is one that is not often explicitly discussed but that, I believe, is implicitly assumed in much of the work on this. And this is that in 1b, for example, the fact

that the answer is understood as conveying the proposition that Claribel left the party at midnight ‘comes for free’. After all, according to SLM, that is exactly what 1b says. To elaborate, let us revisit the dialogue in 1 between our two speakers A and B.

- (1) a. A: Who left the party at midnight?  
 b. B: Claribel.

There is no doubt that anyone listening to this (including speaker A or our eavesdropper C) will conclude that B intends to convey the proposition that Claribel left the party at midnight. And this is automatic if 1b actually MEANS ‘Claribel left the party at midnight’, which in turn follows if linguistically it actually has the form *Claribel left the party at midnight*. I suspect that the fact that 1b seems to have the relevant proposition as its meaning lies behind a good deal of the original intuition underlying SLM.

But does 1b really mean the same as *Claribel left the party at midnight*? Actually, we have no direct way of knowing. What is understood by this utterance obviously varies completely from dialogue to dialogue. This is quite different from the situation with respect to stable truth conditions of normal declarative sentences. Of course, particular conveyed meanings of a sentence can vary from dialogue to dialogue, but stable truth conditions do not.

So it is worth reminding ourselves of the actual empirical fact, which is only that from listening to B’s response in the context of the dialogue in 1, one can conclude that B intends to convey the proposition *Claribel left the party at midnight*. Put differently, assuming that B is knowledgeable and honest, one can conclude that this proposition is true. This is what needs to be explained under any analysis. To say something stronger—for example, that the linguistic expression *Claribel* by itself in 1b maps to the proposition that Claribel left the party at midnight—is not an empirical fact. It is a theory-dependent claim.

So given the actual empirical fact that needs to be explained, SLM has no advantage. Both theories need an account of the fact that our eavesdropper (or any participant in the conversation) arrives at the conclusion that speaker B intends to convey the relevant proposition, and saying that 1b is ‘really’ *Claribel left the party at midnight* is of no help. After all, listeners have no direct access to the silent strikethroughs. Under the SLM analysis, we still have to answer the question of how it is that the listener computes the strikethrough material. In other words, how does the processor arrive at the meaning of material that it does not hear? This question (especially the analogous question in cases of VPE) has been extensively discussed in the processing literature (see e.g. Murphy 1985, Frazier & Clifton 2005, Snider & Runner 2011, and many others), and detailed discussion of those results is beyond the current scope. But the important point is that the SLM hypothesis has no more immediate account of how the answer is understood than does Qu-Ans; it merely shifts the question to one of how the listener infers what the silent material is. (Note too that the account of how it is that the conversational participants arrive at the understanding that they do is rather obvious under Qu-Ans as they can recognize this as a Qu-Ans pair, and the semantics of such pairs gives the relevant proposition.)

A few additional observations are in order regarding the hypothesis (which seems to be implicit in at least many works endorsing SLM) that processing of the answer 1b crucially involves constructing some sort of representation whereby 1b is the full sentence *Claribel left the party at midnight*, and where that representation is then processed. There are two obvious ways to flesh this out. The first is that listeners literally supply a phonological string such as [left the party at midnight] that has already been heard, combine

this with [Claribel], and then process the result in exactly the way they would have had they directly heard the phonological material [Claribel left the party at midnight]. But first, it would be a somewhat surprising discovery that listeners recompute the meaning of a string like [left the party at midnight] when this has already been computed. Plausibility aside, we know that that cannot be right because of the indexical shifts: it is INTERPRETED material (i.e. content) that is supplied and combined with the short answer. But then, the processing story is pretty much just like what the Qu-Ans analysis says: in arriving at the inference that Claribel left the party at midnight, two meanings are combined. (There is a difference between the two accounts: in SLM it is the meaning of a sub-part of the question *t left the party at midnight* that is crucial, whereas in Qu-Ans it is the meaning of the full question that combines with the meaning of the answer.) In sum, the way in which the dialogue in 1 is understood is completely orthogonal to the question of whether the grammar treats the answer as having a full sentence as its representation.

**6. SHORT ANSWERS SUPPORT DIRECT COMPOSITIONALITY.** This section is short. Once again I repeat the logic behind many of the arguments for SLM.

- A. PREMISE: Phenomenon P has to be stated with respect to some property of (nonlocal) chunks of representation.
- B. FACT: We find phenomenon P in fragment answers.
- C. CONCLUSION: Therefore there must be a level of representation at which the fragment answer is surrounded by other material.

As noted several times, DC would lead us to believe that the premise in A is wrong in any case and that for each relevant phenomenon, we need to find alternative accounts that do not refer to nonlocal chunks of representation.

We can now turn the logic of these arguments on their head. For we have seen in §4 that there is evidence that the conclusion in C is incorrect; positing that the short answer is covertly a fuller sentence causes problems in accounting for the phenomena discussed there. I cannot claim to have definitely shown that no SLM analysis can work, but these facts certainly call into question the plausibility of SLM. And to the extent that there is evidence that the conclusion in C is wrong, then, since B is correct, we now know that A must be wrong too. In other words, the DC claim that each phenomenon needs to be accounted for by mechanisms that do not refer to chunks of representation is vindicated by the behavior of short answers to questions.

**7. CONCLUSION.** This article has tried to motivate the conclusion that each of the following four questions has a very short answer: (i) Is there good reason to believe that fragment answers to questions contain silent linguistic material? The short answer: ‘No’. (ii) Is there evidence that short answers indeed do not contain linguistic material, but combine directly with the meaning of the question? The short answer: ‘Yes’. (iii) Is the correct analysis of short answers compatible with direct compositionality? The short answer: ‘Yes’. (iv) Does the correct analysis of short answers actually provide evidence for direct compositional analyses of phenomena such as reflexives, binding of pronouns, and so forth? The short answer: ‘Yes’.

#### REFERENCES

- ANDERBOIS, SCOTT. 2014. The semantics of sluicing: Beyond truth conditions. *Language* 90.887–926. DOI: 10.1353/lan.2014.0110.
- BACH, EMMON. 1979. Control in Montague grammar. *Linguistic Inquiry* 10.515–31. Online: <http://www.jstor.org/stable/4178132>.

- BECK, SIGRID, and HOTZE RULLMANN. 1999. A flexible approach to exhaustivity in questions. *Natural Language Semantics* 7.249–97. DOI: 10.1023/A:1008373224343.
- BOŠKOVIĆ, ŽELKO, and HOWARD LASNIK. 2003. On the distribution of null complementizers. *Linguistic Inquiry* 34.527–46. DOI: 10.1162/002438903322520142.
- CHOMSKY, NOAM. 1981. *Lectures in government and binding*. Dordrecht: Foris.
- CULICOVER, PETER, and RAY JACKENDOFF. 2005. *Simpler syntax*. Oxford: Oxford University Press.
- DOWTY, DAVID. 1982. Grammatical relations and Montague grammar. *The nature of syntactic representation*, ed. by Pauline Jacobson and Geoffrey K. Pullum, 79–130. Dordrecht: Reidel.
- ENGD AHL, ELISABET. 1986. *Constituent questions: The syntax and semantics of questions with special reference to Swedish*. Dordrecht: Reidel.
- ENGD AHL, ELISABET. 1988. Relational interpretation. *Mental representations: The interface between language and reality*, ed. by Ruth Kempson, 63–82. Cambridge: Cambridge University Press.
- FIENGO, ROBERT, and ROBERT MAY. 1994. *Indices and identity*. Cambridge, MA: MIT Press.
- FRAZIER, LYN, and CHARLES CLIFTON, JR. 2005. The syntax-discourse divide: Processing ellipsis. *Syntax* 8.121–74. DOI: 10.1111/j.1467-9612.2005.00077.x.
- GAZDAR, GERALD; EWAN KLEIN; GEOFFREY K. PULLUM; and IVAN SAG. 1985. *Generalized phrase structure grammar*. Oxford: Blackwell, and Cambridge, MA: Harvard University Press.
- GINZBURG, JONATHAN, and IVAN SAG. 2000. *Interrogative investigations: The form, meaning, and use of English interrogatives*. Stanford, CA: CSLI Publications.
- GRIMSHAW, JANE. 1982. Grammatical relations and subcategorization. *Subjects and other subjects: Proceedings of the Harvard conference on grammatical relations*, ed. by Annie Zaenen, 35–55. Bloomington: Indiana University Linguistics Club.
- GROENENDIJK, JEROEN, and MARTIN STOKHOF. 1984. *Studies in the semantics of questions and the pragmatics of answers*. Amsterdam: University of Amsterdam dissertation.
- GROENENDIJK, JEROEN, and MARTIN STOKHOF. 1989. Type shifting rules and the semantics of interrogatives. *Properties, types and meaning, vol. 2: Semantic issues*, ed. by Genaro Chierchia, Barbara Partee, and Raymond Turner, 21–68. Dordrecht: Kluwer.
- HAMBLIN, C. L. 1973. Questions in Montague English. *Foundations of Language* 10.41–53. Online: <http://www.jstor.org/stable/25000703>.
- HANKAMER, JORGE, and IVAN SAG. 1984. Toward a theory of anaphoric processing. *Linguistics and Philosophy* 7.325–45. DOI: 10.1007/BF00627709.
- HEIM, IRENE, and ANGELIKA KRATZER. 1998. *Semantics in generative grammar*. Oxford: Blackwell.
- HELLER, DAPHNA. 2002. On the relation of connectivity and specificational pseudoclefts. *Natural Language Semantics* 10.243–84. DOI: 10.1023/A:1022152007523.
- HIGGINS, ROGER. 1973. *The pseudo-cleft construction in English*. Cambridge, MA: MIT dissertation. [Reprinted, Bloomington: Indiana University Linguistics Club, 1976.]
- JACOBSON, PAULINE. 1994. Binding connectivity in copular sentences. *Proceedings of Semantics and Linguistic Theory (SALT)* 4.161–78. DOI: 10.3765/salt.v4i0.2456.
- JACOBSON, PAULINE. 1999. Towards a variable-free semantics. *Linguistics and Philosophy* 22.117–85. DOI: 10.1023/A:1005464228727.
- JACOBSON, PAULINE. 2000. Paycheck pronouns, Bach-Peters sentences, and variable-free semantics. *Natural Language Semantics* 8.77–155. DOI: 10.1023/A:1026517717879.
- JACOBSON, PAULINE. 2003. Binding without pronouns (and pronouns without binding). *Resource-sensitivity, binding, and anaphora*, ed. by Geert-Jan Kruiff and Richard Oehrle, 57–93. Dordrecht: Kluwer.
- JACOBSON, PAULINE. 2007. Direct compositionality and variable-free semantics: The case of ‘principle B’ effects. *Direct compositionality*, ed. by Chris Barker and Pauline Jacobson, 191–236. Oxford: Oxford University Press.
- JACOBSON, PAULINE. 2009. The short answer. And what it answers. Paper presented at the 16th annual HPSG Conference, Göttingen, Germany.
- KAPLAN, DAVID. 1989. Demonstratives. *Themes from Kaplan*, ed. by Joseph Almog, John Perry, and Howard Wettstein, 481–563. Oxford: Oxford University Press.

- KARTTUNEN, LAURI. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1.3–44. DOI: 10.1007/BF00351935.
- KEENAN, EDWARD L. 1971. Names, quantifiers, and a solution to the sloppy identity problem. *Papers in Linguistics* 4.211–32. DOI: 10.1080/08351817109370257.
- KRIFKA, MANFRED. 2001. Quantifying into question acts. *Natural Language Semantics* 9.1–40. DOI: 10.1023/A:1017903702063.
- KUNO, SUSUMU. 1975. Three perspectives in the functional approach to syntax. *Chicago Linguistic Society (Parasession on functionalism)* 11.276–336.
- LAKOFF, GEORGE. 1971. Presupposition and relative well-formedness. *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*, ed. by Danny Steinberg and Leon Jakobovits, 329–40. Cambridge: Cambridge University Press.
- MAY, ROBERT. 1985. *Logical form: Its structure and derivation*. Cambridge, MA: MIT Press.
- MERCHANT, JASON. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.
- MERCHANT, JASON. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27.661–738. DOI: 10.1007/s10988-005-7378-3.
- MERCHANT, JASON; LYN FRAZIER; CHARLES CLIFTON, JR.; and THOMAS WESKOTT. 2013. Fragment answers to questions: A case of inaudible syntax. *Brevity*, ed. by Laurence Goldstein, 21–35. Oxford: Oxford University Press.
- MONTAGUE, RICHARD. 1970. English as a formal language. *Linguaggi nella Società e nella Tecnica*, ed. by Bruno Visentini, 189–224. Milan: Edizioni di Comunità.
- MORGAN, JERRY L. 1973. Sentence fragments and the notion ‘sentence’. *Issues in linguistics: Papers in honor of Henry and Renée Kahane*, ed. by Braj Kachru, Robert Lees, Yakov Malkiel, Angelina Pietrangeli, and Sol Saporta, 719–51. Urbana: University of Illinois Press.
- MURPHY, GREGORY L. 1985. Processes of understanding anaphora. *Journal of Memory and Language* 24.290–303. DOI: 10.1016/0749-596X(85)90029-4.
- PARTEE, BARBARA, and EMMON BACH. 1981. Quantification, pronouns, and VP anaphora. *Formal methods in the study of language: Proceedings of the 8th Amsterdam Colloquium*, ed. by Jeroen Groenendijk, Theo M. V. Janssen, and Martin Stokhof, 445–81. Amsterdam: University of Amsterdam Mathematical Center.
- POLLARD, CARL, and IVAN A. SAG. 1992. Anaphors in English and the scope of binding theory. *Linguistic Inquiry* 23.261–303. Online: <http://www.jstor.org/stable/4178768>.
- POLLARD, CARL, and IVAN A. SAG. 1994. *Head driven phrase structure grammar*. Stanford, CA: CSLI Publications, and Chicago: University of Chicago Press.
- POSTAL, PAUL. 1964. Underlying and superficial linguistic structure. *Harvard Educational Review* 34.246–66. DOI: 10.17763/haer.34.2.jxp438774xp522k7.
- ROOTH, MATS. 1985. *Association with focus*. Amherst: University of Massachusetts dissertation.
- ROOTH, MATS. 1992. Ellipsis redundancy and reduction redundancy. *Proceedings of the Stuttgart ellipsis workshop*, ed. by Steven Berman and Arild Hestvik. Heidelberg: IBM.
- ROSENBAUM, PETER. 1967. *The grammar of English predicate complement constructions*. Cambridge, MA: MIT Press.
- ROSS, JOHN ROBERT. 1985. The source of pseudocleft sentences. Handout from New York University colloquium talk.
- SAG, IVAN. 1976. *Deletion and logical form*. Cambridge, MA: MIT dissertation.
- SAG, IVAN, and JOANNA NYKIEL. 2011. Remarks on sluicing. *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar (HPSG 2011)*, 188–208. Online: <http://csli-publications.stanford.edu/HPSG/2011/sag-nykiel.pdf>.
- SCHLENKER, PHILIPPE. 2003. Clausal equations (a note on the connectivity problem). *Natural Language Semantics* 21.157–214. DOI: 10.1023/A:1021843427276.
- SCHULZ, KATRIN, and ROBERT VAN ROOIJ. 2006. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy* 29.205–50. DOI: 10.1007/s10988-005-3760-4.
- SHARVIT, YAEL. 1999. Connectivity in specificational sentences. *Natural Language Semantics* 7.299–339. DOI: 10.1023/A:1008390623435.
- SNIDER, NEAL, and JEFFREY T. RUNNER. 2011. Phonological structure if reactivated in VP ellipsis: Evidence from eye movements. Paper presented at the annual meeting of the Linguistic Society of America, Pittsburgh, PA.

- STANTON, ROBERT. 2005. In defense of non-sentential assertion. *Semantics versus pragmatics*, ed. by Zoltán Gendler Szabó, 383–457. Oxford: Oxford University Press.
- STEEDMAN, MARK. 1987. Combinatory grammars and parasitic gaps. *Natural Language and Linguistic Theory* 5.403–39. DOI: 10.1007/BF00134555.
- STEEDMAN, MARK. 1996. *Surface structure and interpretation*. Cambridge, MA: MIT Press.
- SZABOLCSI, ANNA. 1987. Bound variables in syntax (are there any?). *Proceedings of the 6th Amsterdam Colloquium*, ed. by Jeroen Groenendijk, Frank Veltman, and Martin Stokhof, 331–53. Amsterdam: University of Amsterdam Mathematical Center.
- WARD, GREGORY, and JULIA HIRSCHBERG. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61.747–76. DOI: 10.2307/414489.
- ZRIBI-HERTZ, ANNE. 1989. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language* 65.695–727. DOI: 10.2307/414931.

Department of Cognitive, Linguistic,  
and Psychological Sciences  
Brown University  
[pauline\_jacobson@brown.edu]

[Received 30 November 2012;  
revision invited 4 June 2013;  
revision received 4 July 2014;  
accepted 7 July 2015]