1

Timing Puzzles in Anaphora and Interpretation

Andrew Barss

1 Introduction

This chapter has two main goals. First, I will present a model of the syntax–semantics mapping which differs from the standard models in derivational syntax by building the logico-semantic representation for a sentence incrementally, in parallel with the syntactic derivation. On this view, there is no designated level of representation Logical Form (LF). Rather, LF is a derivative concept, simply referring to the endpoint of the dual syntactic and semantic derivations. The second goal of the chapter is to anchor a theory of anaphoric reconstruction effects within this model, and to characterize reconstruction as a particular instance of the formation of syntactically constrained semantic relations in the middle of an ongoing derivation.

Development of this particular approach to the determination of anaphoric relations is critically tied to a strong interpretation of the notion "time point in a derivation," specifically with respect to the Earliness Principle (Pesetsky 1989, Diesing 1992). The core idea to be explored is that anaphoric dependencies (and other syntactically constrained semantic relations) must be formed as early as possible in a derivation. Data examined will include some of the well-known reconstruction effects with anaphors and R-expressions, as well as other anaphoric relations including weak crossover effects with bound pronominals, and the interaction of these effects with ellipsis and the A/A′-movement distinction. This analysis of anaphoric relations will then be set in a broader context in sections 3 and 4, in which I discuss several puzzles in the syntax–semantics mapping which do not receive an easy account within the standard model. I will argue that these puzzles may be resolved under a model in which Earliness applies to all semantic relations, and semantic interpretations are built up in parallel with syntactic operations. The general discussion will also examine closely what is formally entailed by the informal notion "as soon as possible." Section 5 concludes the discussion, with consideration of the interplay of morphological licensing and semantic Earliness.

2 The Parallel Model Sketched

The issue of how the syntax–semantics mapping is achieved is one that has been central to generative grammar since its inception. Within movement-based models, a key controversy has been which level of syntactic representation feeds the computation of semantics.[1] Several distinct answers have been given. In the Standard Theory, semantics was computed off D-structure, and so movement transformations were necessarily meaning-preserving. In REST and P&P, a certain division of semantic labor between syntactic levels was achieved by adopting two leading ideas: first, that LF was the sole structural input to the semantic component, and second, that LF retained certain properties of D-structure (and S-structure) forced by the interplay of the θ-criterion, the projection principle, and trace theory. Essentially, argument–structure relations were established at D-structure, and remained unaffected by subsequent movement operations. Scope relations between quantifiers, negation, and the like were determined by movement (overt or covert) operations, and formally instantiated at LF by relative c-command.

Within the Minimalist Program, the level D-structure has been dispensed with, in favor of a derivational syntax that incrementally builds syntactic structure. All lexical insertion is limited to the overt (pre-SPELLOUT) segment of the derivation, and so we may view the overt syntax as incrementally building up argument–structure relations, as new predicates and arguments are added. To take a simple example, consider the sentence in (1), with its derivation in (3):

(1) John saw Mary kiss Fred.
(2) Numeration = {John, Mary, Fred, kiss, saw, I^0_{+Past}, I^0_{def}}
(3) a. [kiss Fred]
    b. [I^0_{def} [kiss Fred]]

[1] I use the singular term "level" here deliberately, since almost all work in derivational grammar has supposed that only a single level of syntactic structure is directly visible to the rules of semantic interpretation – LF in the GB and MP models, D-structure in the Standard Theory model.

c. [Mary [I^0_{def} kiss Fred]]
d. [saw [Mary [I^0_{def} kiss Fred]]]
e. [I^0_{+Past} [saw [Mary [I^0_{def} kiss Fred]]]]
f. [John [I^0_{+Past} [saw [Mary [I^0_{def} kiss Fred]]]]]

Since there is no single "all at once" establishment of argument relations in this model, one might suppose two approaches to the point(s) in the derivation at which argument–structure relations are determined:

(4) Predicate–argument relations are established in toto immediately after LF.
(5) Each predicate–argument relation is determined as the argument is Merged.

(4) has been the standard model since the groundbreaking work of Chomsky (1977, 1981), May (1977, 1985), and Higginbotham (1980, 1983a). However, every formal implementation of (4) presumes that the actual logico-semantic representation for a given LF representation is built up incrementally, by the successive application of rules of interpretation working within local subparts of the tree. Compositional semantic systems exist which can work from the top of a tree down, or the bottom of a tree up, but *any* formal interpretation system must compute a sentence's meaning in a sequence of steps, each contributing an additional piece of semantic structure to what was previously computed. In other words, even if we suppose (with the standard view) that there is but one point at which the entire syntactic structure is handed off to the semantic component, there will be a derivation within the semantic system itself. Assuming a derivational syntax, then, we can take as a bedrock principle that there is a syntactic derivation and a semantic derivation, and (via the compositionality thesis) that the two are intimately tied. The syntactic derivation builds a layered phrase marker in a series of local-structure-affecting steps, and the semantic derivation computes the contribution to meaning of each of these local subregions of the phrase marker.

As an alternative to the standard model, I wish to explore some of the ramifications (and, I will argue, advantages) of a generalized version of (5). The remainder of this section will sketch the basics, and the subsequent sections will fine tune it. I shall refer to the general model as the "parallel semantics model." On this view, the syntactic derivation in (3) is accompanied by a parallel computation of semantic relations, given in (6). Each row in the table reflects a step in the parallel derivation, with the syntactic structure in the left column, semantic structure on the right. Boldface shows new material added. Labeling is kept at a minimum for clarity:

(6) Overt derivation for (1), on parallel-semantics model:[2]

Num = {John, Mary, Fred, kiss, saw, I^0_{+Past}, I^0_{def}}	
Syntactic operation	*Semantic relation established*
[kiss Fred]	**theme(Fred, e)**
[I^0_{def} [kiss Fred]**]**	**∃e** theme(Fred, e)
[Mary [I^0_{def} kiss Fred]]**]**	∃e theme(Fred, e) **& Agent(Mary, e)**
[saw [Mary [I^0_{def} kiss Fred]]]**]**	**Percept (**[∃e theme(Fred, e) & Agent(Mary, e)], **e′)**
[I^0_{+Past} [saw [Mary [I^0_{def} kiss Fred]]]]**]**	**(∃e′, e′<u)** Percept ([∃e theme(Fred, e) & Agent(Mary, e)], e′)
[John [I^0_{+Past} [saw [Mary [I^0_{def} kiss Fred]]]]]**]**	(∃e′, e′<u) Percept ([∃e theme(Fred, e) & Agent(Mary, e)], e′) **& Experiencer(John, e′)**

Following the analysis of perceptual reports developed by Higginbotham (1983b), the one remaining step to map the last line in (6) onto a complete semantic representation is to scope-assign the embedded event description, as in (7):

(7)

Input	
[John [I^0_{+Past} [saw [Mary [I^0_{def} kiss Fred]]]]]	(∃e′, e′<u) Percept ([∃e theme(Fred, e) & Agent(Mary, e)], e′) **& Experiencer(John, e′)**
Covert syntactic operation	*Semantic relation established*
[Mary [I^0_{def} kiss Fred]]$_1$ [John [I^0_{+Past} [saw **e**$_1$]]]	**(∃e: theme(Fred, e) & Agent(Mary, e))** (∃e′<u) Percept (**e**, e′) & Experiencer(John, e′)

[2] Some comments on notation and theoretical assumptions: I am assuming here the event semantics developed by Davidson (1967), Higginbotham (1983a, 1983b, 1985), Kratzer (1996), and Parsons (1990). Each event-type verb introduces, as one of its arguments, a variable over events, which is typically existentially bound by a functional head in the VP-external inflectional complex (here represented as I^0, since the particular details of the inflectional complex do not concern us here). I am assuming here a bare-bones theory of tense, on which semantic tense orders the event time with respect to utterance time **u**. See Zagona (this volume) and Stowell (1998a, 1998b) for more sophisticated theories of tense relations.

Such a model differs from the single-input models which occur in the existing literature. There is no one syntactic structure which feeds the computation of semantic relations. Rather, as the syntactic structure is incrementally enlarged, the semantic representation is successively amended, adding new material to the previously computed material, via rules which effectively interpret the newly added syntactic material (a Merged argument is assigned a thematic role, a Merged I^0 binds its event variable and orders it with respect to utterance time, etc.). The default is that all of the previously computed semantic information is carried over to the next stage, just as the syntactic default in Minimalist syntax is to completely carry over the previously computed phrase marker in its entirety to subsequent stages (Chomsky's 1992, 1995a Extension Condition).

What advantages does this model have over the standard conceptions of the syntax-to-semantics mapping? Conceptually, it accords a degree of symmetry to the syntax and the semantics, but are there empirical arguments in its favor? In the rest of this chapter I will explore a series of case studies which support the incremental-semantics model, which will also serve as opportunities to more fully specify the operations of and constraints on the parallel model.

3 Anaphoric Timing Puzzles

Within the Minimalist framework, all binding relations apply at LF (Chomsky 1995a, 1995b). However, there are a number of paradigms which seemingly conflict with this proposal.

Higginbotham (1988) observes a grammaticality distinction in the following pair (on the bound-variable construal of the pronoun), which he analyzes as a weak crossover (WCO) effect:

(8) [Which musician]$_1$ [t$_1$ played [which piece that you asked him$_1$ to play]]?
(9) ?* [Which piece that you asked him$_1$ to play]$_2$ did [which musician]$_1$ play t$_2$?

The puzzle emerges when we realize that at LF, they have equivalent structures (linear order of the two operators aside):

Each thematic role is a two-place relation between an argument and the event, with e.g. Agent(Mary, e) to be read as "Mary is the Agent of event e." Finally, I assume (following Higginbotham 1983a) that bare perception-verb complements denote a specific event description, and are transported covertly to a scope position via QR (see (7)).

(10) [which piece that you asked him to play]$_2$ [which musician]$_1$ [t_1 play t_2] ?

The exact nature of the WCO effect is controversial and has been given several different formal treatments, but I assume the principle in (11), adapted from Higginbotham (1983a) and Barss (1995):

(11) *Weak Crossover Principle:* If a pronoun is to be coindexed with a *wh*-operator, then it (the pronoun) must be c-commanded by a trace dependent on the operator.

(11) is violated in the standard WCO example (13):

(12) [Which man$_1$ [t_1 kissed [his$_1$ wife]]]?
"[for which x, x a man] [x kiss x's wife]?"
(13) * [Which man$_1$ did [[his$_1$ wife] kiss t_1]]?
"[for which x, x a man] [x's wife kiss x]?"

Assuming LF *wh*-movement, the WCO principle is also violated in the LF representation (10). Higginbotham's discussion centers on this point, and concludes that the relevant condition must apply at S-structure. The pronoun is properly licensed in the S-structure (8), but not in (9).

In a framework lacking a formal level S-structure, some other account obviously must be provided. It turns out, however, that Higginbotham's conclusion can be quite easily maintained. The abstract generalization is that if the configuration described in (11) is met *prior* to LF, bound-variable anaphora is permitted. What is central to Higginbotham's argument is that the licensing configuration exists in (9) and is subsequently obliterated when the *wh*-phrase raises at LF. This, apparently, is sufficient to license anaphora. As a provisional statement we may amend (11) to (14):

(14) *Weak Crossover Principle:* If a pronoun is to be coindexed with a *wh*-operator, then it (the pronoun) must be c-commanded by a trace dependent on the operator at some derivational stage (not necessarily at LF).

This is reminiscent of Belletti and Rizzi's (1988) approach to Condition A. They analyze the *amuse* class of psych verbs as raising predicates, and account for the grammaticality of (15, 16) by proposing that Condition A is an "anywhere" condition, which has to be satisfied at some point in a derivation but can be violated at earlier or later stages:

(15) [Pictures of themselves] amuse the men
(16) [The men believe [pictures of themselves] to amuse me

Both (14) and Condition A are "positive" licensing conditions: if they are satisfied, the anaphoric relation is permitted. Condition C, on the other hand, is a negative licensing condition: if the configuration described is met, the anaphora is disallowed. It has been known for some time that Condition C is apparently sensitive to S-structure configurations. The original observation by Brody (1979) is that covert *wh*-movement cannot bleed Condition C:

(17) Joe wonders [which book that Mary checked out of the library]$_2$ [she put t$_2$ [where$_3$]]
(18) * Joe wonders [where$_3$ [she$_1$ put [which book that Mary$_1$ checked out of the library]$_2$ t$_3$]]

After LF *wh*-movement:

(19) Joe wonders [[[which book that Mary$_1$ checked out of the library]$_2$ where$_3$ [she$_1$ put t$_2$ t$_3$]

Brody concludes that Condition C applies at S-structure, since, were it to apply solely at LF, (18) would be grammatical.

This argument was revisited by Chomsky (1992, 1995a), who proposes as an alternative that there is no covert (phrasal) *wh*-movement. In (18), the *wh*-phrase remains in situ, with the [+*wh*] determiner (or just the [+*wh*] feature) raising to mark the scope of the interrogative. If adopted, this proposal then means that Brody's paradigm is consistent with an LF-only application of Condition C.

However, the same problem exists in constructions which are not amenable to the "no covert movement" thesis. Consider the pairs below:

(20) Mary$_1$ expects me to bring more chairs than her$_1$ husband did.
(21) Mary$_1$ expects me to bring more chairs than her$_1$ husband does.

(22) * She$_1$ expects me to bring more chairs than Mary$_1$'s husband did.
(23) * She$_1$ expects me to bring more chairs than Mary$_1$'s husband does.

Condition C suffices to rule out (22, 23), since the pronoun c-commands the R-expression with which it is coindexed. However, this is only true in the overt, i.e. S-structure, representation, since the comparative AP *more chairs than . . .* contains an ellipsis site. This is particularly problematic in (23), since the ellipsis takes wide scope, and includes the matrix verb. Following Fiengo and May's (1994) extensive defense of the LF-raising theory of ellipsis resolution, I assume that in both (21) and (23) the embedded AP raises into the matrix clause, out of the matrix VP (which is then reconstructed into the ellipsis site (boldface material indicates the reconstructed VP, following Fiengo and May's notation):

(24) [more chairs than her_1 husband PRES **expects me to bring $\mathbf{e_2}$** $]_2$ Mary_1 expects me to bring e_2 (= LF of (21))

(25) [more chairs than Mary_1's husband PRES **expects me to bring $\mathbf{e_2}$** $]_2$ she_1 expects me to bring e_2 (= LF of (23))

The wide-scope raising of NP bleeds the Condition C configuration, and so Condition C cannot apply only at LF (contra Chomsky's 1995a argument, reviewed above).

What I would like to suggest, as a reconciliation of these puzzles with the Minimalist thesis that S-structure is not a formal level of representation, is the following:

(26) *Earliness of anaphoric relations:* Form and filter anaphoric dependencies as early as possible in a derivation.

I assume, following Evans (1977), Higginbotham (1983a), Fiengo and May (1994), and Barss (1995), that an anaphoric dependency is formally encoded as an asymmetric relation between a dependent item and its antecedent. Several notations have been employed in the literature. I here write D(A, B) to express the dependency of an expression A upon its antecedent B. Dependencies are formal constructs, akin to movement chains, and so one can think of them as relations which are explicitly formed in the course of a derivation.

We also need to specify "as soon as possible" with some care. For a simple coreference dependency, let us suppose that the dependency can be formed (and filtered) as soon as both NPs are merged into the same tree, i.e. when they are both present in the same "active buffer" of the syntactic derivation. The Condition C puzzles in (17) vs. (18), and (21) vs. (23), now follow directly, without reference to S-structure. In both (18) and (23) Condition C is violated at an earlier step of the derivation than LF. In pre-Minimalist terms this led to the conclusion that Condition C applies at S-structure, but we retain the same result with (26). In (23), the pre-SPELLOUT representation meets the requirements for dependency formation: both the pronoun and the R-expression have been Merged into the same phrase marker, and so the dependency D(*she*$_1$, *Mary*$_1$) is formed, and immediately filtered out as ungrammatical by Condition C. The LF geometric relation between the two NPs is irrelevant. Similarly, we retain the essence of Brody's analysis of (18). D(*she*$_1$, *Mary*$_1$) is formed in the overt part of the derivation, and again immediately filtered out by Condition C.

Formally, this does not occur at exactly the immediately pre-SPELLOUT stage of the derivation (S-structure, in GB terms). The interpretation suggested for "as soon as possible" actually forces the dependency to be formed immediately after Merger of the higher NP. So, in e.g. (23), the dependency is formed immediately after (27) is computed by Merge:

(27) She$_1$ expects me to bring more chairs than Mary$_1$'s husband does.

Let us now return to the WCO puzzle that began this section. In contrast to simple coreference between two overt NPs, bound-variable anaphora between an operator and a pronoun requires three participating elements: the pronoun; the operator; and a trace dependent on the operator. The licensing principle (14) makes reference to all three, and to their configurational relationship. Consequently, the dependency of the pronoun on the operator cannot be formed until all three elements are present in the same phrase marker. Since the trace occurs as a result of the movement of the operator, it follows that a bound-variable anaphoric dependency cannot be formed until the operator moves.

For example (8), the dependency D(*him*$_1$, *which musician*$_1$) will be formed immediately after *wh*-movement of *which musician*, that is, at the final overt stage (i.e. at S-structure, in pre-Minimalist terms). Following (26), this dependency will be immediately filtered, and licensed as grammatical, by (14). Subsequent movement of the lower *wh*-phrase in the LF component – which "breaks" the licensing relation – is irrelevant, since the dependency has already been licensed. Thus we derive the results of Higginbotham's stipulation that (14) applies at S-structure.

For the WCO violation (9), repeated here, the dependency cannot be formed until the in-situ *wh which man* moves (this time in the covert component), leaving trace. (I assume here that the choice of which *wh*-phrase is moved overtly is essentially free; see Pesetsky 1987, Reinhart 1995, Barss 2000 for discussion.) When it does, the representation (10) is formed:

(9) ?* [Which piece that you asked him$_1$ to play]$_2$ did [which musician]$_1$ play t$_2$?

(10) [which piece that you asked him to play]$_2$ [which musician]$_1$ [t$_1$ play t$_2$] ?

By (26), the dependency D(*him*$_1$, *which musician*$_1$) is formed immediately after (10) is computed. This violates (14), and hence WCO arises, since t_1 does not c-command *him*.

The proposal in (26) derives the classical results of binding and anaphoric licensing conditions seeming to apply uniquely to S-structure, without appeal to this level of representation, and so is fully consistent with a major thesis of Minimalist syntax. It is important to note that, if this line of argument is correct, there is no one point in the grammar at which a specific anaphoric licensing condition applies. Where in a particular derivation a dependency is formed, and filtered, is highly variable, contingent on the sequence of Merge operations, the exact type of dependency (coreference, bound-variable anaphora), and the number of elements which must be present in a phrase marker for the dependency to exist at all.

4 Generalized Earliness

The proposal (26) is inspired by (and indeed is simply a special application of) Pesetsky's (1989) Earliness Principle, a general condition requiring that all constraints on representation be satisfied as early as possible within a derivation. Pesetsky's paper presents the principle as a preference scale of levels of representation:[3]

(28) *Earliness Principle:* Satisfy principles as early as possible on the hierarchy of levels: (DS >) SS > LF > LP. (Pesetsky's (13))

The version of Earliness I discuss here is different from Pesetsky's. It is more Minimalist, in that it dispenses with references to levels of representation; and it is somewhat more fine-grained, in that it forces the application of anaphoric filters at a very specific point in a derivation, rather than dividing the derivation up into covert, overt, and language-particular blocks, as Pesetsky's does. Nonetheless, I believe (26), and the extensions of it I present below, to be fully consistent with the conceptual core of Pesetsky's principle.

Anaphora is one case, among many, of a linguistic phenomenon which involves the participation of syntax, morphology, and rules of semantic

[3] LP is a level of representation reserved for language-particular rules, like English *do*-support, which plays a prominent role in Pesetsky's theory.

interpretation. We can view each anaphoric principle as specifying, for a particular type of linguistic formative α (a pronoun, a trace, an anaphor, etc.), the type of syntactic configuration required to support a particular type of interpretation involving α and another referential element. At the beginning of this chapter, I sketched out an incremental model of the syntax–semantics mapping in which syntactic and semantic derivations occur in parallel, with each step of the syntactic derivation potentially feeding its cohort in the semantic derivation. The core example was the successive computation of argument relations, which I suggested could plausibly be viewed as being entered in the semantic representation immediately upon Merger of the argument with the (projection of) the θ-marking head. What I would like to now pursue is the natural extension of (5) and (26) in this model. The principle would be informally stated as follows:

(29) *Earliness of Semantic Interpretation (**ESI**):* Compute the semantic interpretation of elements of a phrase marker as early as possible in a derivation.

We continue to adopt the interpretation of "as early as possible" defended in section 3, whereby all the elements mentioned in an interpretive principle must be present in the same phrase marker before the interpretation is effected. With argument structure, this constraint is met when the argument Merges. With anaphoric relations, as discussed above, the constraint is met when the dependent element, the antecedent, and any other necessary participant (e.g. the trace in bound-variable construal of pronouns) are all subparts of the same phrase marker.

In the subsections below, I would like to present a series of other cases which substantiate this more general proposal.

4.1 Adverbial scope

Consider the following examples:

(30) John said that [yesterday Mary had left].
(31) John said yesterday that [Mary had left].
(32) # John will say that [tomorrow Mary had left].
(33) John will say tomorrow that [Mary had left].

It is an obvious fact that the temporal adverbial in (30) and (31) is interpreted with respect to the clause in which it overtly occurs (the embedded clause in

(30), the matrix clause in (31)). It is also an obvious fact that (32) is unacceptable, and certainly cannot mean what (33) does.

But if semantic interpretation is computed directly off LF representations, as the classic theory states, a problem arises. Suppose, in each case, the adverbial moves covertly, so that (30–2) become the LFs (30′–32′):

(30′) John said yesterday that [Mary had left].
(31′) John said that [yesterday Mary had left].
(32′) John will say tomorrow that [Mary had left].

(I assume, with Lasnik and Saito 1984, 1992, that movement may occur without leaving trace, so long as the output representation is interpretable.) If this movement occurred, and if it is the resulting LFs which are semantically interpreted, then (30) and (31) should be ambiguous, and synonymous, and (32) should be fully acceptable. Clearly, either such movement has to be blocked (in the syntax), or the classical theory of LF and interpretation has to be modified.

Let me note first that the ESI readily accounts for these cases. Several theories of adverbial interpretation exist (for discussion, see Chierchia and McConnell-Ginet 2000, Kratzer 1996, Larson 1988, and Larson and Segal 1995), some viewing a temporal adverbial as an argument of V, some as a modifier (of VP or a temporal projection). Regardless of which option we choose, the crucial ingredients necessary for the adverbial to be interpreted are the adverbial itself and the head or projection it modifies or is an argument of. And, following the arguments of section 3, the two must be Merged into the same phrase marker prior to the computation of the adverbial's contribution to meaning. All of these preconditions are met in (30, 31, 32). As a result, the covert movement which on hypothesis gives rise to the LFs (30′–32′) is semantically vacuous. The adverb might move, but its meaning is already computed.

One might object that there is already a perfectly good syntactic reason why the mapping from (30–2) to (30′–32′) is invalid, namely that the movement is unnecessary, hence blocked by Economy (that is, the derivation which does nothing to (30) is shorter, and hence more Economical, than the one which maps (30) onto (30′)). However, recent work on Economy principles has seen a need to limit the syntactic Economy conditions to regulating choices among semantically equivalent derivations (see particularly Fox 1995). Since the derivations under discussion shift interpretation, they should be permitted under Economy. I thus conclude that the paradigm above, a problem for the classical model of the syntax–semantics mapping, supports the alternative parallel model.

4.2 Bittner's puzzle

In her 1992 "movement and scope," Maria Bittner observes that overt head raising has a null semantic effect, in contrast to phrasal movement, which does have an effect semantically. Bittner illustrates this with the relative scope of negation and modals:

(34) fordi [Jakob ikke kan komme]
because J. not can come
(35) [CP Jakob1 [C kan2] [IP t1 ikke [t2 komme]]]
Jakob can not come
available: not [possible [J. comes]]
unavailable: possible [not [J. comes]]

In both examples (independent of the presence or absence of verb raising to C^0), negation obligatorily takes wide scope over the modal.

Although the modal *kan* in (35) has raised to COMP, out of the c-command domain of negation, it continues to obligatorily take scope within negation (as in the non-movement example (34)). So, raising a scopal verb over another scope-bearing item does not "reverse" the scope of the two items.

Bittner's theoretical characterization is to propose that, at LF, the raised head (*kan* in (35)) must lower back into its pre-movement position, thus effectively undoing movement before semantic interpretation. The explanation I offer here is essentially the converse of Bittner's: the relative scope of the heads is computed as soon as they are co-Merged into the phrase marker, and hence subsequent movement (overt here) is semantically vacuous, since it follows interpretation.

A paradigm rather similar to Bittner's can be constructed in English, given certain assumptions about the saturation of the event position carried by the verb. The discussion in 4.3 is partially based on Barss and Basilico (1994).

4.3 Generalization of Bittner's paradigm: English verb raising, negation scope, and events

Consider the meaning of (36):

(36) John has not left

Crucial to the interpretation of (36) is wide scope of the negation over the event and the perfective aspectual auxiliary, so that (36) is true iff it is not the case that there is a completed event of leaving by John (within the contextually salient period of time).

I assume, following Belletti (1990), and many others, that the phrase structure of (36) is such that NegP dominates TP, which dominates the auxiliary verb's projections, which dominate the thematic VP. Thus I assume that the derivation of (36) is essentially as in (37):

(37) a. [John left]
b. $[_{\mathrm{VP}}$ has $[_{\mathrm{VP}}$ John left]]
c. $[\mathrm{T}^{0}_{\mathrm{pres}}.$ $[_{\mathrm{VP}}$ has $[_{\mathrm{VP}}$ John left]]]
d. [has$_1$ +$\mathrm{T}^{0}_{\mathrm{pres}}$. $[_{\mathrm{VP}}$ t_1 $[_{\mathrm{VP}}$ John left]]]
e. $[_{\mathrm{NegP}}$ not [has$_1$ +$\mathrm{T}^{0}_{\mathrm{pres}}$. $[_{\mathrm{VP}}$ t_1 $[_{\mathrm{VP}}$ John left]]]]
f. [AGR $[_{\mathrm{NegP}}$ not [has$_1$ +$\mathrm{T}^{0}_{\mathrm{pres}}$. $[_{\mathrm{VP}}$ t_1 $[_{\mathrm{VP}}$ John left]]]]]
g. [[has$_1$ +$\mathrm{T}^{0}_{\mathrm{pres}}]_2$+AGR $[_{\mathrm{NegP}}$ not [t_2 $[_{\mathrm{VP}}$ t_1 $[_{\mathrm{VP}}$ John left]]]]]
h. [John$_3$ [[has$_1$ +$\mathrm{T}^{0}_{\mathrm{pres}}]_2$+AGR $[_{\mathrm{NegP}}$ not [t_2 $[_{\mathrm{VP}}$ t_1 $[_{\mathrm{VP}}$ t_3 left]]]]]]

Steps (a), (b), (c), (e), and (f) reflect successive Merge of the verb and its argument, the auxiliary, Tense, negation, and AGR. Steps (d) and (g) reflect cyclic verb raising to T and AGR, and (h) reflects subject raising to the highest functional-head SPEC position.

By the ESI hypothesis, each of the Merge steps (a–c) and (e) should result in immediate attempted interpretation of the material Merged. (I assume AGR to be semantically vacuous, hence step (f) results in no additional semantic computation.) If the material Merged on these steps can be fully interpreted when Merged, then the Move steps (d), (g), and (h) should be semantically vacuous.

I further assume, following Davidson (1967), Higginbotham (1983a, 1983b), and Parsons (1990), that event-class verbs carry an argument position for events, that in tensed clauses this e-position is discharged by existential quantification, and that it is Tense which serves as the existential quantifier. I further suppose, following Parsons, that the contribution of perfective *have* is essentially to assert that the resultant state (Parson's R-state) of the event holds (in whatever time frame corresponds to the verb's tense morphology, e.g. Present in the examples discussed here). Thus (38) comes out as (39):

(38) John has left
(39) (∃e) [leaving(e) & Agent(John, e) & hold(e's R-state, now)]

The final conjunct in (39) is provided by the auxiliary *have* in Present tense.

We obviously want the interpretation of (36) to be the logical negation of (39), given as (40):

(40) ¬ (∃e) [leaving(e) & Agent(John, e) & hold(e's R-state, now)]

Computing this interpretation in parallel with the syntactic derivation in (37) is unproblematic, since the ordering of semantic material in (40) closely mirrors the Merged order of elements in (37). Here is the parallel derivation, with boldface indicating new syntactic or semantic material:

(41)

Num = {John, left, T^0_{PRES}, not, has}	
Syntactic derivation	*Logico-semantic derivation*
[John left]	**[leaving(e) & Agent(John, e)]**
[$_{VP}$ has [$_{VP}$ John left]]	[leaving(e) & Agent(John, e) **& hold(e's R-state, now)]**
[T^0_{pres} [$_{VP}$ has [$_{VP}$ John left]]]	**(∃e)** [leaving(e) & Agent(John, e) & hold(e's R-state, now)]
[has$_1$ +T^0_{pres} [$_{VP}$ t_1 [$_{VP}$ John left]]]	vacuous; no change
[$_{NegP}$ not [has$_1$ + T^0_{pres} [$_{VP}$ t_1 [$_{VP}$ John left]]]]	**¬** (∃e) [leaving(e) & Agent(John, e) & hold(e's R-state, now)]
[AGR [$_{NegP}$ not [has$_1$ + T^0_{pres} [$_{VP}$ t_1 [$_{VP}$ John left]]]]]	vacuous
[[has$_1$ +T^0_{pres}]$_2$+AGR [$_{NegP}$ not [t_2 [$_{VP}$ t_1 [$_{VP}$ John left]]]]]	vacuous
[John$_3$ [[has$_1$ +T^0_{pres}]$_2$+AGR [$_{NegP}$ not [t_2 [$_{VP}$ t_1 [$_{VP}$ t_3 left]]]]]]	vacuous

On the parallel model, two things of importance occur. First, lexical material and syntactic structure is interpreted incrementally, and as soon as possible, in the unmarked case upon Merge. Second, movement of heads, and A-movement of the subject, force no change in the semantic representation. As a result, we derive exactly the correct interpretation. Negation, the last (highest) non-vacuous head Merged, is interpreted with logical scope over the event quantifier and the verbs' argument structure.

Now consider what happens on the classical model, in which the final syntactic cell in (41) – the LF of the sentence – is semantically interpreted, and the previous syntactic derivational stages are invisible to the semantic interpretation system. Notice that overt head movement of T^0 and perfective *have* has moved these items structurally above negation. Mutatis mutandis, we would expect the interpretation to be that in (42):

(42) $(\exists e)$ hold(e's R-state, now) $\neg$ [leaving(e) & Agent(John, e)]

(42) is true just in case there is some event whose resultant state holds now (i.e. the event is completely done), and it is something other than a leaving by John, for example a smoking of a cigar by Smith. This is the wrong result, since there will always be such events, independent of whether John stayed or left. Thus the classical theory, in its simplest form, produces the wrong result, just as in Bittner's original examples.

Bittner's original account suffices to account for the English paradigm presented here: type considerations force LF lowering of a raised head. However, set in a broader context, it would appear that both paradigms are subcases of a more general phenomenon: lexical material, and syntactic structure, introduced in the course of a syntactic derivation are semantically interpreted as soon as they possibly can be, in the limiting case on the derivational step in which that lexical material is Merged. Heads have a multiplicity of semantic function. Some are logical operators (negation, modals), some introduce thematic roles and assign them to referential or propositional argument phrases (verbs), some function as variable-binding quantifiers (Tense), and some introduce abstract predication of events (auxiliary *have*). Yet across this panoply of head types, one property seems apparently constant: each of these heads can be interpreted immediately upon Merge. Given the constraint encoded as the ESI Principle, these heads thus must be interpreted upon Merge, and so head movement is (derivatively) semantically vacuous.

There is actually at least one well-attested exception to this. On Larson's (1988) theory of VPs, a ditransitive verb cannot assign its highest (i.e. external) θ-role when it is Merged. This is the result of the interaction of Larson's theory of VP-projections, particularly the dual assumptions that phrase structure is binary branching and that maximal projections host at most one specifier. As a consequence, the verb *give* in (43) is not fully interpreted in its Merged position. Rather, V raises into a higher V position, at which point it can θ-mark its highest argument. Although this proposal was something of an anomaly in the GB model, since all θ-marking occurred at D-structure in that framework, Larson's proposal fits naturally with the general framework presented here:

(43) Nancy gave Mary a book

(44) Parallel derivation:

Syntactic derivation	*Semantic derivation*
[$_{V'}$ gave [Mary]]	**giving(e) & Goal(Mary, e)**
[$_{VPa}$ [$_{V'}$ gave [Mary]] **[a book]]**	giving(e) & Goal(Mary, e) **& Theme(a book, e)**
[$_{V'}$ [gave]$_1$[$_{VPa}$ [$_{V'}$ e$_1$ [Mary]] [a book]]**]**	no change
[$_{VPb}$ Nancy [$_{V'}$ [gave]$_1$[$_{VPa}$ [$_{V'}$ e$_1$ [Mary]] [a book]]]**]**	giving(e) & Goal(Mary, e) & Theme(a book, e) **& Agent(Nancy, e)**
[$_{TP}$ T^0 [$_{VPb}$ Nancy [$_{V'}$ [gave]$_1$[$_{VPa}$ [$_{V'}$ **e$_1$** [Mary]] [a book]]]]]	**($\exists$e<u)** giving(e) & Goal(Mary, e) & Theme(a book, e) & Agent(Nancy, e)
[AGR [$_{TP}$ T^0 [$_{VPb}$ Nancy [$_{V'}$ [gave]$_1$ [$_{VPa}$ [$_{V'}$ e$_1$ [Mary]] [a book]]]]]]	no change
[$_{AgrP}$ Nancy$_2$ [AGR [$_{TP}$ T^0 [$_{VPb}$ e$_2$ [$_{V'}$ [gave]$_1$[$_{VPa}$ [$_{V'}$ e$_1$ [Mary]] [a book]]]]]]]	no change

4.4 Deriving strict cyclicity for wh-*movement*

Considerable discussion has been devoted to a problem that arises in the formal explanation of the *wh*-island effect. Descriptively, a *wh*-phrase cannot grammatically be moved from a [+*wh*] clause, as in the standard example (45):

(45) * What$_1$ did John wonder [where$_2$ Mary bought t$_1$ t$_2$]]

Chomsky (1977), in a landmark discussion, observed that there are two derivations which must be precluded. On the first derivation, *where* moves to the embedded [SPEC, CP], and on a subsequent step *what* moves in one fell swoop to the matrix [SPEC, CP], as in (46):

(46) What$_1$ did John wonder [where$_2$ Mary bought t$_1$ t$_2$]]

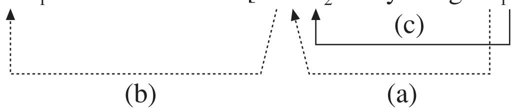
(a)

(b)

Any formulation of Subjacency will account for the ungrammaticality of this derivation, since *what* moves over too large a region in its sole movement.

The second derivation is one in which *what* moves to the matrix [SPEC, CP] via the lower [SPEC, CP] (steps (a) and (b) below), and *where* subsequently moves to the lower [SPEC, CP] (step (c) below):

(47) What$_1$ did John wonder [where$_2$ Mary bought t$_1$ t$_2$]]



This latter derivation does not violate Subjacency, if that principle is stated as a constraint on movement, since no movement step in (47) crosses an island. Chomsky (1977) proposes a separate principle, Superiority, to block this alternative derivation. Superiority prevents movement on a lower cycle once a higher cycle has been affected by movement. Chomsky (1992) reinterprets Superiority as the Extension Condition, requiring that derivations only affect material at the top of the currently active phrase marker. As a result, the movement of *where* in step (c) is blocked, because it does not reposition *where* at the very top of the phrase marker. Extension is a somewhat problematic condition, as Chomsky himself notes. Its range is limited to substitution operations (not adjunction operations), and apparently limited to the overt component of the grammar.

The ESI Principle proposed above provides an alternative explanation for the ill-formedness of the derivation in (47), one not requiring additional Superiority or the Extension Condition. Observe that in this derivation, *what* moves through a [+*wh*] SPEC, and subsequently moves to a higher one. *Wh*-phrases have one type of scope position: the specifier of a [+*wh*] C^0. Once *what* is moved to the lower SPEC, CP, it is fully interpretable: it binds a variable, and is in a proper scope position for an interrogative. Hence, by ESI, it *must* be interpreted in that scope position, and is blocked from subsequent movement on the matrix cycle. Whereas on Chomsky's (1977) and (1992) accounts, it is the final movement of *where* which dooms this derivation, on the account proposed here it is the second movement of *what*.

5 Morphological Licensing and Semantic Earliness

In this final section, I would like to tie together a timing puzzle explored in detail by Fiengo and May (1994), and what appears to be a conflict in the literature concerning reconstruction from A-positions. This will lead to a conjecture, partially presaged by Fiengo and May's theory of reconstruction, which points to a striking interplay of SRI and morphological licensing.

Fiengo and May observe that there is a grammaticality contrast in the following pair, considered under coreference between the pronoun and the R-expression:[4]

(48) Mary introduced him_1 to everyone that $John_1$ wanted her to
(49) * Mary introduced him_1 to everyone that $John_1$ wanted her to introduce him_1 to

Fiengo and May argue that S-structure application of Condition C fails to distinguish the examples, since in both cases the highest occurrence of the pronoun c-commands the name in the overt representation. In order to properly resolve the ellipsis in (48), the embedded QP will have to raise in the LF component to a position in the matrix clause:

(50) [everyone that $John_1$ wanted her **to introduce $\mathbf{him_1}$ to $\mathbf{e_2}$**] [Mary introduced him_1 to e_2]

Fiengo and May claim that (48) gives evidence for QR bleeding Condition C, and for LF application of Condition C. The problem is that other examples indicate that Condition C must apparently apply at S-structure. These include (49), if wide-scope QR is a free option, and cases like my (23), in which Condition C must apply before QR (wide-scope QR is forced by ellipsis in (23), hence the debate over whether wide-scope QR is a free option is irrelevant). Thus a puzzle: why does Condition C apply before QR (i.e. at S-structure) in cases like my (23), while it applies only after QR (i.e. only at LF) for cases like (48)?

Fiengo and May observe a striking fact about the feed–bleed relation occurring in examples with antecedent-contained deletion (ACD). QR (forced to resolve ellipsis) bleeds Condition C just in case the elided material contains an occurrence of the index involved in the anaphoric dependency. They argue that binding theory can only apply to indices which are fully projected. Ellipsis is a case where the lexical material is not projected until the identity condition on ellipsis is satisfied. For cases of ACD, like the unpronounced VP in (48), the identity condition is only satisfied after QR. Thus, the index 1 is not fully licensed until after QR, and so, by Fiengo and May's proposed constraint, Condition C cannot apply until LF in examples like (48/50).

By contrast, in my (23), and in the non-ellipsis case (49), the relevant index is fully licensed at S-structure. Hence, by their proposal, Condition C can apply

[4] My (48) is Fiengo and May's example (95a), p. 294, and my (49) is their (167).

to the index at S-structure, and will rule out coreference in both examples. Thus, the paradoxical feed–bleed relation between QR and Condition C in cases of ACD is resolved. Condition C applies as soon as it can, but under the restriction that every occurrence of the index involved in the referential dependency must be fully projected for it to apply.

To slightly reinterpret their proposal in terms of the SRI I have advanced above, let us adopt (51) as a constraint on dependency formation:

(51) An anaphoric dependency D(A, B) cannot be formed (or filtered) until both A and B are morphologically licensed.

In my discussion below, "morphological licensing" will cover two related properties: lexical projection and feature checking.

Chomsky (1995b) argues that reconstruction from derived A-positions is impossible. This argument is made partially on the basis of example (52):

(52) John expected him to seem to me [t to be intelligent]

Chomsky's argument is as follows. Suppose reconstruction (by which Chomsky means literal LF lowering, followed by LF application of the binding conditions) can occur, mapping the S-structure (52) onto the LF (53):

(53) John expected Ø to seem to me [him to be intelligent] (LF for (52), if reconstruction applies)

At this point, the lowered pronoun *him* is sufficiently far below *John* that Condition B should rule the sentence grammatical. But coreference is indeed blocked in (52). For Chomsky, this must mean that reconstruction is not allowed from A-positions, since it is axiomatic in his framework that binding conditions only apply at LF.

The framework developed in this chapter seems to give a more adequate account of this example. We simply note that, in the course of the derivation, even assuming such lowering to be possible, (52) derivationally precedes (53). Hence, the dependency D(him_1, $John_1$) must be formed at the earlier stage, and filtered there by the binding conditions. (52) will, by ESI, be filtered out as a Condition B violation.

However, this explanation has a flaw. In the current framework anaphoric dependencies are formed in the course of a derivation, at the earliest possible point at which the dependent item and its antecedent are constituents of the same tree. Going back to the example above, (52) does not in fact represent the earliest derivational stage at which *him* and *John* have both been Merged into the same phrase marker. Consider an even earlier point in the derivation, before overt raising of the pronoun:

(54) John expected Ø to seem to me [him to be intelligent]

(54) is of course isomorphic to (53), and the same problem arises as for LF-reconstruction theory. Somehow, dependency formation must be delayed until after the pronoun raises.

My slight modification of Fiengo and May's proposal, given in (51), provides the reason for this delay. Prior to the raising of the pronoun, the pronoun's chain is not fully licensed, since the Case of the pronoun has yet to be checked. It is only after raising, i.e. in (52), that the pronoun's chain, and hence the index borne by it, is morphologically licensed. The morphological precondition on dependency formation, together with the ESI Principle, conspires to force (52) to be the exact syntactic structure in which dependency formation must occur.

This proposal has advantages over Chomsky's. First, scope facts suggest that lowering in the LF component from A-positions should in general be permitted (May 1977, 1985), so the assertion that the lowering in the mapping from (52) to (53) is impossible is somewhat dubious. Second, the literature is replete with examples of anaphor reconstruction from derived subject positions, including Belletti and Rizzi's (1988) classic demonstration. (See Barss (2001) for review and discussion.) Consider, for example, the psych-verb construction in (55), and the cyclic A-movement example (56):

(55) [each other$_1$'s houses]$_2$ [[impressed t$_2$] [the architects]$_1$ immensely]
(56) [each other$_1$'s houses]$_2$ [seem [t$_2$ to have [[impressed t$_2$] [the architects]$_1$ immensely]]

For further detailed discussion of the A/A′ distinction and reconstruction, see Ueyama (this volume).

6 Conclusion

In this chapter, I have proposed that the computation of semantic information carried by a phrase marker be computed in the course of a derivation, incrementally building up the semantics in a fashion closely tied to the internal operations of the derivational syntax. (For a more detailed and formal development of this model, see Barss (forthcoming).) This model resolves a series of timing puzzles in coreference and bound-variable anaphora, and covers related puzzles in the interpretation of negation and other head-level operators.

This general parallelism of syntactic and semantic computation, I have argued, is subject to two fundamental constraints. The first is Earliness of Semantic Interpretation, forcing the semantic interpretation of material to follow its introduction into the phrase marker as rapidly as possible. We have shown that the effects of this principle vary considerably, depending on the nature of the semantic relation involved. The second constraint, derived from Fiengo and May's work on ellipsis resolution, requires that anaphoric dependencies be formed only after their participant chains have been fully licensed by the morphosyntactic processes of the grammar. Fuller development of this system, and particularly exploration of the generality of the second constraint beyond anaphoric dependencies, awaits further research.

Acknowledgments

I am grateful to students and colleagues at the University of Arizona for feedback on the material presented here, parts of which have been presented in seminars and a colloquium in 1997 and 1999. I am particularly indebted to David Basilico, Jim McCloskey, Mario Montalbetti, and Simin Karimi for discussion and inspiration.