

A semantic constraint on *wh*-movement
Extended events and extraction from *in order* clauses

Robert Truswell

The existence of well-formed cases of *wh*-movement out of adjuncts constitutes a challenge to current locality theories. However, the fact that such extractions are not universally well-formed leaves an empirical hole in any theory which simply removes the adjunct condition without putting anything in its place. I show that the patterns of legitimate extraction out of adjuncts can be described in terms of an independently motivated theory of the internal structure of events. This means that our syntactic theory of locality needs to be supplemented by independent semantic well-formedness conditions on *wh*-movement.

1. Introduction

This paper attempts to sketch a program for dealing with classes of observed exceptions to an otherwise quite robust generalisation, which has come to be known as the *adjunct condition*. This condition has its roots in the seminal work of Ross (1967).¹ In its purest form, it can be stated as in (1).

- (1) *Wh*-dependencies cannot be formed across adjunct boundaries.

As with all the other constraints which Ross discusses, and no doubt partially as a consequence of the formal tools available in 1967, the adjunct condition is entirely syntactic. By this, I mean that the condition is defined by sole reference to nodes in a phrase structure tree and relations among those nodes. In the work on locality theory stimulated by Ross' dissertation, this syntactocentrism has remained. Huang's (1982) Condition on Extraction Domain, for example, represents a genuine milestone in our understanding of locality effects, and it retains the syntactocentrism of Ross' approach, by defining extraction domains in terms of the syntactic relation of proper government. For our purposes in this paper, though, the CED moves beyond Ross'

¹In actual fact, only tensed adjuncts automatically fall under Ross' definition of *island*, which makes crucial reference to S nodes occurring outside a chain of complementation relations. In the subsequent years, though, adjuncts in general have widely been assumed to be strong islands.

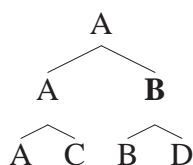
approach in two respects. Firstly, it extends Ross' prohibition against extraction out of tensed adjuncts to an explicit ban against extraction from *any* adjunct. Secondly, it unifies the subject and the adjunct conditions. As neither subjects nor adjuncts are properly governed, they are both strong islands for extraction.

In recent years, since Chomsky (1995) questioned the validity of government relations, several papers, such as Uriagereka (1999); Johnson (2002) and Zwart (2007), have attempted to derive CED effects from other considerations.² Although these three papers start from different sets of concerns, they end up with strikingly similar conclusions. These are summarised in (2).

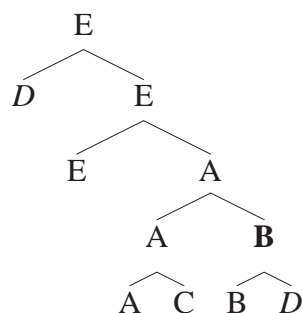
(2) A non-projecting phrase is a strong island if its sister is a phrase.

Pictorially, this means that in the tree in (3a) (where linear order should be taken to be irrelevant), the higher node labelled B is an island, and so a dependency formed across it, such as movement of D in (3b), is ungrammatical.

(3) a.



b. *



Of course, this is not the only conceivable syntactocentric position. For example, a position gaining some currency in HPSG (see, for example, Levine & Sag 2003), is that adjuncts, and possibly also subjects, are simply not islands, with the degradation of many cases of extraction from adjuncts coming from largely unspecified extragrammatical factors. Although I will eventually adopt something not a million miles from this position, such an approach in isolation could be seen as somewhat unsatisfying, given that in most cases, extraction from adjuncts is quite simply, and quite categorically, bad. Nothing can rescue cases like the following.³

- (4) a. *What did you come here [because Mary wanted to talk to you about *e*]?
 b. *What did John go home [after Mary said *e*]?
 c. *What was little Jimmy playing in the sandpit [screaming about *e*]?

²Van de Koot (2004) gives a related, but subtly different approach. As this paper is no less syntactocentric, it remains prone to the sort of criticism to be developed below.

³I stick to extraction of complements from within adjuncts throughout this paper. Needless to say, extraction of subjects or adjuncts from within adjuncts is even worse.

According to most popular syntactic frameworks, the grammar works in black and white. Chomsky (1995), for example, only gives two possible outcomes of a syntactic derivation. It crashes, or it converges.⁴ On the other hand, the extragrammatical factors appealed to by Levine & Sag may well be somewhat more gradient. The examples in (4) look pretty categorical, but are they representative? It turns out, in fact, that extraction from adjuncts is unquestionably bad in a great many cases, but that it is not bad across the board. This paper discusses two classes of adjuncts which allow subextraction in at least some cases, namely *bare present participial adjuncts*, as illustrated in (5) and discussed in more detail in Truswell (2007a); and *in order* clauses, as in (6).

- (5) a. What did John drive Mary crazy [whistling *e*]?
 b. What did John cut himself [carving *e*]?
 (6) a. What are you working so hard [in order to achieve *e*]?
 b. What did you come in [to talk to us about *e* today]?

This paper aims to show that these two groups of exceptions to the adjunct condition form a natural class. However, that class is described, not in the syntactic terms which have dominated discussion of extraction from adjuncts, but rather in semantic terms relating to the internal structure of events. This leads to the claim that there is a semantic condition on \bar{A} -constructions, related to event structure, in addition to the well-studied syntactic constraints.

The rest of the paper is structured as follows: Section 2 discusses the distribution of legitimate extractions from these two classes of adjuncts in more detail, and introduces the core semantic condition on *wh*-movement to be discussed in this paper. Section 3 then develops a novel theory of the internal structure of events, motivated by a series of experiments reported in Wolff (2003). Finally, section 4 relates this theory of event structure to the locality data given in section 2, and offers a brief comparison with syntactic alternatives.

2. The problem

The existence of grammatical apparent cases of extraction from adjuncts has been noted in the literature since at least the early days of Government and Binding theory.⁵ Perhaps Cinque (1990:ch.3) provides the fullest discussion of such cases. The core of Cinque's very detailed analysis is that the gap within an adjunct island is never actually a trace of movement, but a null pronominal (*pro*) which comes to be \bar{A} -bound by a sufficiently local operator. It then transpires that the locality conditions on \bar{A} -dependencies across adjunct boundaries reduce to locality conditions on (\bar{A} -)binding. If, in certain cases, binding of *pro* contained within an adjunct by an operator outside that adjunct is possible, then an \bar{A} -dependency with many superficial charac-

⁴The third possibility discussed by Chomsky, convergence as gibberish, is not syntactically distinct from regular convergence, but rather distinct at the interfaces.

⁵The most famous such case is, of course, the *parasitic gap* phenomenon illustrated in (i), but I put these aside here on account of their quite distinct properties.

(i) What did John file *t* [without reading *e*]?

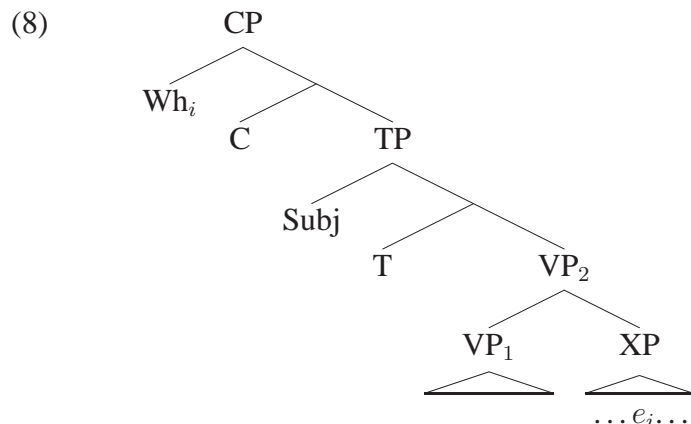
teristics of movement is possible across the adjunct boundary.

Cinque gives us a clear and plausible way to account for many facts concerning extraction out of adjuncts, which I cannot do justice to here. To give just one example of the empirical value of his approach, though, it derives the fact that only DPs are ever legitimately extracted from adjuncts, as noted by Chomsky (1982:72-3) in connection with the examples in (7). This follows automatically if the gaps within adjuncts are actually *pro*, and *pro* is a DP.

- (7) a. The man that I went to England without speaking to *e*
 b. *The man to whom I went to England without speaking *e*

However, one area where we gain no insight from Cinque's theory concerns the distribution of the types of adjuncts out of which we can form such dependencies in the first place. Some adjunct maximal projections apparently do not constitute barriers to binding, whereas others apparently are barriers. But, in either case, it could just as easily have been the other way round.

More problematically for an approach such as Cinque's, it appears that the acceptability of extraction from an adjunct is determined partially by the nature of the VP to which it is adjoined. This is quite unexpected on an approach based on barriers, or intervention on the path between the head and foot of the chain, as the VP to which the adjunct attaches is not part of that path. Schematically, in a structure such as (8), we often find that the nature of the material in VP₁, to which XP is adjoined, can affect the acceptability of the dependency between Wh_{*i*} and *e_i*, despite the fact that VP₁ is not actually on the path between the two members of the chain and so cannot plausibly count as a barrier to the formation of that dependency.



The following examples of extraction from bare present participial adjuncts illustrate this claim. (9) shows that the adjunct *complaining about the weather* can felicitously modify both of the matrix VPs in question. However, (10) shows that extraction is only possible in one of the two cases in question. On the assumption that the two sentences are syntactically identical apart from the content of the matrix VP, the distinction between (10a) and (10b) is a mystery on either a post-Cinque or a post-CED approach.

- (9) a. John drove Mary crazy [complaining about the weather].
 b. John does his work [complaining about the weather].

- (10) a. What did John drive Mary crazy [complaining about *e*]?
 b. *What does John do his work [complaining about *e*]?

The question I want to address here, then, is this: Why are some adjunct boundaries more *transparent* than others? That is, why can dependencies cross some adjunct boundaries, but not others?

The answer I will propose to that question deviates from the syntactocentric view which has persisted from Ross through to Uriagereka, Johnson, and Zwart. In a nutshell, I believe that nothing narrowly syntactic distinguishes a legitimate case of extraction of a DP from an adjunct by *wh*-movement, from a degraded example of the same: both are syntactically quite legitimate. Instead, they are distinguished in terms of a second set of structural constraints, concerning event semantics. The condition I propose is the following.

(11) **Events define locality domains for *wh*-movement**

A *wh*-chain is legitimate only if the denotation of the minimal constituent containing the head and foot of the chain asserts the existence of a single event.

In an earlier paper, Truswell (2007a), I showed that a condition along the lines of (11) could make sense of the pattern in (9)–(10) if we adopt a decompositional view of Vendler's (1957) aspectual classes, as in, for example, Dowty (1979) and Parsons (1990). Vendler claimed that verb phrases could be divided into four classes, the *accomplishments*, *achievements*, *activities* and *states*, on the basis of two intersecting families of diagnostic tests. For example, accomplishments and activities readily allow use of the progressive, as shown in (12a-b), whereas this is at least degraded for achievements and states, as in (12c-d).

- (12) a. I am running a mile (drawing a circle, building a house, . . .).
 b. I am running (writing, working, . . .).
 c. *I am spotting the plane (appearing, blinking, . . .).
 d. *I am knowing the answer (loving you, understanding antisymmetry, . . .).

On the other hand, accomplishments and achievements reject *wh*-questions of the form *for how long* . . ., as in (13a) and (13c), while these are quite acceptable with activities (13b) and states (13d).

- (13) a. #For how long did he run a mile (draw a circle, build a house, . . .)?
 b. For how long did he run (write, work, . . .)?
 c. #For how long did you spot the plane (appear, blink, . . .)?
 d. For how long did you know the answer (love me, understand antisymmetry, . . .)?

Dowty and Parsons give a way of deriving this pattern. If we assume, following Fodor (1970) among others, that a single verb phrase describes (or denotes a property of) a single event, then one interpretation of Dowty's and Parsons' approach is that such an event can be decomposed into a maximum of two salient subevents, which I will call a *process* (conceived of as durative and lacking a culmination) which immediately precedes and directly causes a pointlike *culmination*, where *direct causation* may be defined as follows.

- (14) An event e_1 *directly causes* an event e_2 iff:
- a. e_1 causes e_2 ;
 - b. There is no event e_3 at the same level of granularity⁶ as e_1 or e_2 such that e_1 causes e_3 and e_3 causes e_2 .

The different aspectual classes can be defined, to a first approximation, in terms of the presence or absence of those two subevents, as follows, where an arrow represents direct causation. Vendler's tests can then be explained if we assume that progressives can only be formed from VPs whose denotation includes a process component, and *for how long* questions can only be formed from VPs whose denotation *does not* include a culmination.

- (15) a. **Accomplishment** (e.g. *build a house, draw a picture*): Process \rightarrow culmination.
 b. **Achievement** (e.g. *arrive, notice the commotion*): (Process \rightarrow) culmination.⁷
 c. **Activity** (e.g. *work, dance*): Process.
 d. **State** (e.g. *know French, love Mary*): \emptyset .

In the terms of this approach, (10a) can be distinguished from (10b) in terms of aspectual class membership. In both cases, the adjunct event *complain about what* is an activity. In (10b), the matrix event *do his work* is also an activity. In (10a), on the other hand, the matrix event *drive Mary crazy* is an accomplishment. Moreover, it is an accomplishment in which the nature of the preparatory process is underspecified. I claim in Truswell (2007a) that an example like (10a) can meet condition (11), on the decompositional approach pioneered by Dowty and Parsons, if the process described in the adjunct VP is identified as the preparatory process in the accomplishment-denoting matrix VP. Accordingly, it is clear that, in (10a), the only possible interpretation is one in which Mary becomes crazy as a direct result of John's whistling.

As for the ungrammatical (10b), we see that in this case, both the matrix VP and the adjunct VP describe fully specified processes. Therefore, unlike the previous case, identification of these two events is not possible. This is for the simple reason that a complaining event is not an event of doing work: the descriptive content attached to the two event variables is not compatible, and so the identification cannot take place. Without such identification, condition (11) cannot be satisfied and extraction is impossible.

As one of the central elements in this explanation is the requirement that the matrix process subevent be underspecified, we predict that parallel cases of extraction from a bare present participial adjunct modifying an accomplishment with a fully specified preparatory process component should be ungrammatical. This prediction is borne out by comparing the two examples in (16).

- (16) a. John wrote the cheque [complaining about what a waste of money it was].
 b. *What did John write the cheque [complaining about e]?

Although *write the cheque* is an accomplishment VP, its preparatory process is clearly specified to consist of writing. The complaining event denoted by the adjunct VP therefore cannot be

⁶See in particular Bittner (1999) on this notion.

⁷See Truswell (2007a,b) for discussion of the presence of preparatory processes in achievements. I ignore the issue here.

identified as the preparatory process of the cheque-writing event, and (11) is violated. This derives the unacceptability of (16b). We see, then, that a decompositional event semantics gives us a handle on the otherwise perplexing pattern of extractions in (9)–(10) and (16).⁸

The main aim of this paper is to expand this account to cover the case of *in order* clauses, a class of subject-oriented purpose clauses delimited by Jones (1991).⁹ As (17) shows, extraction of object NPs from these adjuncts is quite free, and unrestricted by the sort of aspectual considerations which regulate extraction from bare present participial adjuncts.

- (17) a. What are you working so hard [in order to achieve *e*]?
 b. Who did John travel to England [to make a sculpture of *e*]?
 c. Whose attention is John jumping up and down [in order to attract *e*]?
 d. What did you tap your nose [in order to signal to Mary *e*]?
 e. What did you come in [to talk to us about *e* today]?

The question, as ever, is why it should be the case that extraction from *these* adjuncts is so much freer than extraction from other classes.¹⁰ I will approach this question in two stages. Firstly, in section 3, I define a more substantial, recursive, structure for event composition, motivated on grounds independent of \bar{A} locality. Then, in section 4, I return to the issue of locality. It will be shown that, although the more powerful event composition system sketched in section 3 does not lose any empirical adequacy with respect to data such as (9)–(10) above, it follows as a natural conclusion of this theory of event structure, in tandem with (11), that \bar{A} -dependencies across *in order* clause boundaries are essentially unrestricted, while \bar{A} -dependencies across bare present participial adjunct boundaries remain sensitive to aspectual class. Needless to say, nothing of the sort follows on any syntactocentric approach that I can conceive of.

3. Extended events

The purpose of this section is to define a more expansive class of event structures than those discussed above with respect to Vendler's aspectual classes, and give some reasons to assume

⁸This presentation could give the impression that extraction from bare present participial adjuncts is only possible when they modify accomplishment VPs. In fact, this is not the case. Extraction is also possible in some cases when they modify achievements, as in (i), and activities, as in (ii).

- (i) a. What did John die [whistling *e*]?
 b. What did John come back [thinking about *e*]?
 (ii) a. What does John wander around [thinking about *e*]?
 b. Which chair did John eat his breakfast [sitting on *e*]?

Different treatments of these are sketched in Truswell (2007a) and Truswell (2007b), but I wish to abstract away from these cases in this paper.

⁹As will be clear from the examples in (17), the actual words *in order* are not a defining characteristic of *in order* clauses. The word *to* in isolation heading an adjunct is frequently sufficient to indicate an *in order* clause. The defining characteristics of *in order* clauses are the subject-orientation and the 'purpose' semantics.

¹⁰I do not wish to give the impression that extraction from just these two classes of adjunct is possible. In fact, a wide range of adjunct types allow extraction, the most well-known perhaps being the participial adjuncts introduced by a preposition, as in (7a) and the more acceptable of the examples in footnote 15. Such examples have been noted by Chomsky (1982) and Cinque (1990), among others.

that these larger events are still, nonetheless, single events.

I take it for granted that, despite the presence of two subevents in the decompositions of accomplishments and achievements given in (15), it is still legitimate to consider any of the aspectual classes listed there as single events. That is, an accomplishment such as *build a house* is readily considered as a single event, even if it consists of two major subevents, a building process and the coming into existence of a house. Similarly, an achievement such as *arrive* is a single event despite perhaps being considered (with Pustejovsky 1991 and Higginbotham 2000, among others) as being composed of a preparatory process (roughly, a journey) leading to a culmination of being at a destination. Call such events, consisting maximally of a process and a culmination, although either may be absent, *core events*.

Of course, the existence of such part–whole relations among events is not a new idea. Bach (1986) was perhaps the first paper to propose it explicitly, extending Link's (1983) lattice-theoretical account of plural and mass reference in nominals to the domain of events. Bach's structure is general enough to allow quite free composition of smaller events (*subevents*) into larger events (*macroevents*). We have no reason *a priori* to expect the composition of smaller events into larger events to stop at the level of the core event, then. Instead, the question of whether we can further compose multiple core events into larger, *extended events* is best decided empirically.

To approach this question from an empirical point of view, though, we need some diagnostic criteria, or at least some heuristics, for deciding where the boundaries between events lie. To be sure, the stuff that happens all around us does not come neatly packaged into discrete chunks that we can call events. However, we have quite robust intuitions that some things are, and some things aren't, single events (although the grey area is admittedly quite large). Just as I claimed above that building a house is a single event despite its being decomposable into two subevents, it seems quite reasonable to state that reading a book for a while and then walking around in circles is *not* a single event, despite the presence, once again, of two salient subevents. How can we make those intuitions sharper, though?

In this section, I want to report some experimental evidence, presented in Wolff (2003), concerning the size of events. Wolff uses two tests to determine when a portion of stuff that happens is perceived as a single event. The first of these tests couldn't possibly be any simpler. Wolff asked the subjects participating in his experiments to report the number of events that they perceived in an animation. As a less direct measure of the same variable, Wolff also asked the subjects to choose between two alternative descriptions of the same animation, one containing a lexical causative and one containing a periphrastic causative.¹¹ The rationale underpinning this latter test stems from a hypothesis due to Fodor (1970), which can be summed up as in (18).

(18) **Fodor's generalisation:**

A single verb phrase can only describe a single event.

Now, if we consider an animation showing a woman acting in such a way that some smoke rising

¹¹A *lexical* causative is one contained within a single VP, where the causal interpretation and the result state are both inherent to the lexical semantics of the verb. A *periphrastic* causative is a biclausal structure, where the subordinating verb, such as *make* or *cause*, expresses little more than a sort of general causation, and the embedded clause describes the result state that arises.

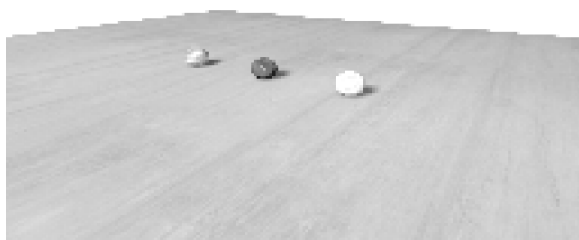
from a nearby ashtray disperses, then we may in principle conceive of this as either a single event, or two separate events, consisting of the woman's actions and the smoke dispersing, respectively. However, the number of events we perceive will influence the options available to us for describing that event, in the light of Fodor's generalisation. If we perceive the animation as depicting a single event, then we may either describe it using a lexical causative, as in (19a), or a periphrastic causative, as in (19b) or (19c). This is because Fodor's generalisation only holds as a one-way implication: just because a single event *can* be described by a single verb phrase, nothing in (18) prevents such an event from being described by multiple VPs.

- (19) a. The woman dispersed the smoke.
 b. The woman made the smoke disperse.
 c. The woman caused the smoke to disperse.

If, on the other hand, we perceive the animation as depicting two separate events, then we cannot describe it using a lexical causative, as to do so would contradict Fodor's generalisation. Only periphrastic options such as (19b-c) are available to describe an animation which is taken to show multiple events, then. In this way, the descriptions of a given animation that subjects produce or accept can cast some light on their perception of events within that animation.

Wolff used these tests in several experiments, each consisting of a series of pairs of animations, differing in some minimal way. In the first experiment, the animations consisted of three marbles in a row, as in (20), from Wolff (2003:16). In the animations, the first marble, A, rolled into the second marble, B, which caused it to roll into the third marble, C, which, in turn, rolled away.

(20)



In the terms used in the discussion of (15) in section 2, then, there are two relations of direct causation holding in such an animation (between A and B, and between B and C), and one relation of indirect causation (between A and C). Given that the taxonomy of core events given in (15) made specific reference to direct causation, we would predict that subjects would report seeing only one event involving marbles A and B, but two events involving marbles A and C. Accordingly, either lexical or periphrastic causatives would be available to describe the interaction of marbles A and B, for example as in (21), but only a periphrastic causative would be available to describe the interaction of marbles A and C, as in (22).

- (21) a. Marble A moved marble B.
 b. Marble A caused marble B to move.

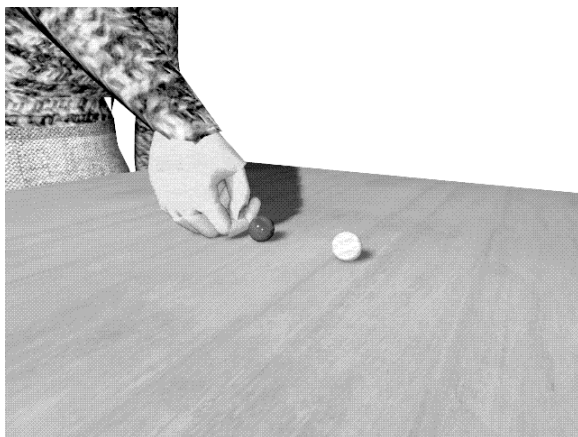
- (22) a. *Marble A moved marble C.
 b. Marble A caused marble C to move.

Indeed, this is what Wolff found. To a highly statistically significant extent, subjects reported an instance of direct causation as a single event, and an instance of indirect causation as two events. Moreover, there was a statistically significant preference for periphrastic causative forms in the description of the indirect causation case.

As it stands, this looks like strong evidence in favour of the position that the core events described in (15) constitute the upper limit for forming macroevents out of smaller subevents. No larger structure involving indirect causation was regularly considered as a single event. However, this finding can be contrasted with the results of Wolff's second experiment.

The setup for the second experiment was identical to that of the first, with the exception that the first marble in the sequence was replaced by a man's hand, as in (23), from Wolff (2003:16).

(23)



Whereas previously, subjects had seen only inanimate objects entering into causal relations, then, there was now an animate instigator for the series of causally related events. This had a striking effect on the perceived event structures of the examples. Specifically, whereas the interaction between marbles A and C in experiment 1 was generally reported to consist of two events, the interaction between the hand and marble C was reported to consist of a single event. Accordingly, both lexical and periphrastic causatives like the following were available to describe the animation.

- (24) a. The man moved marble C.
 b. The man made marble C move.

Given that the only relevant distinction between the animations in the two experiments concerns the animacy of the initiator of the first event, it is reasonable to conclude that the surprising distinction between (22) and (24) is related in some way to animacy. Wolff's third experiment (the last one which I will report here) sheds some light on the nature of this relation.

In the third experiment, two ingredients were kept from experiment 2. Subjects were shown animations consisting of three causally related events, the first of which was initiated by a human. However, in this experiment, the animations were paired in such a way that the human

clearly initiated the chain of causally related events intentionally in one member of the pair, but only accidentally in the other animation. For example, in one animation, a woman is seen wafting her hands, which creates a draught, which in turn causes smoke rising from a nearby ashtray to disperse. In a second animation, on the other hand, a woman walks down a corridor and passes an ashtray from which smoke is rising. The draught she creates by walking causes the smoke to disperse. Stills from these two animations, from Wolff (2003:22), are given in (25).

(25) a.



b.



Although there are arguably three subevents in each case (the woman's actions, the draught, and the dispersal of the smoke), only in the former case can the woman reasonably be said to have deliberately acted in such a way as to make the smoke disperse. Strikingly, the now-familiar pattern was repeated here too. To a highly significant extent, subjects reported perceiving the intentional animations as depicting a single event, and the unintentional animations as depicting two events. Moreover, although both animations were describable with a periphrastic causative such as (26b), only the intentional animation was regularly described by a lexical causative like (26a).

- (26) a. The woman dispersed the smoke.
 b. The woman caused the smoke to disperse.

If the only things which counted as single events were the core event structures described in (15), this result would constitute something of a challenge to Fodor's generalisation (18). According to that generalisation, a single VP, such as *disperse the smoke* in (26a), can only describe a single event. Moreover, the event structures listed in (15) consist maximally of a process and a culmination. However, the animation described by (26a) consists of *two* processes and a culmination. The hand-wafting causes a draught to blow, which, in turn, causes the dispersal of the smoke. Clearly, we cannot simultaneously keep Fodor's generalisation, the foregoing analysis of the animation described above, and the notion that (15) lists the only available types of macroevent formation from subevents, without contradicting ourselves.

As a way out of this, I propose to significantly expand the set of available types of event formation, based on the hypothesis that the difference between the perceptions of the two smoke-dispersal animations described above rests on the notion of intentionality.¹² When the woman was intentionally acting in such a way as to bring about the final event in the causal chain, subjects perceived a single event. When this was not something she intended, subjects perceived two events. In order to formally capture this, I propose a second level of event structure, defined in terms of groupings of the core events described in (15). The definition I propose is the following:¹³

- (27) An extended event consists of a sequence e_1, \dots, e_n of core events, such that:
- a. e_1 occurred and is agentive;
 - b. The agent of e_1 intends e_n to occur;
 - c. For every e_k , $1 \leq k < n$, either e_k causes e_{k+1} or the agent of e_1 believes that the occurrence e_k will enable him to bring e_{k+1} .

Notably, this definition brings an element of recursion into the characterisation of the class of possible subevent structures. This is in parallel to the notions of *forward-* and *backward-chaining* used by Steedman (2002) to characterise our means of plan formation by either working forward towards a goal state from our present state of affairs, or working backward from a goal state in the hope of constructing a plan connecting the present state and the goal state. A well-formed plan, put into action, would meet exactly the conditions described in (27). In a very real sense, then, plans can always be construed as single (extended) events. Space constraints prevent me from defending this position here, but several pieces of linguistic evidence pointing to the validity of this claim are presented in Truswell (2007b).

It should be clear how this definition of extended events allows us to draw the line between the two smoke-dispersal animations above, while maintaining Fodor's generalisation. In the single-event animation (25a), the woman *intended* her actions to bring about the dispersal of the

¹²This presentation differs somewhat from that in Wolff (2003), most notably in not referring to force dynamics, in the sense of Talmy (1988), as Wolff does. However, as far as I can tell, both explanations are equally capable of capturing the data that Wolff presents.

¹³A reviewer raises the issue of the validity of considering such a chain of events as a single event, given that it is possible for the initial subevent in such a chain (the agent's action) to occur without the intended result also occurring. As far as I can tell, this is unproblematic if we adopt a modal accessibility relation among worlds, parallel to the relations yielding sets of belief-worlds and desire-worlds in Heim (1992), for example. Although a dissociation of the initial and final subevents is possible in the real world, the two both necessarily occur in an accessible world corresponding to the agent's plan.

smoke, whereas this was not a plan of hers in the two-event case, (25b). Schematically, we may represent the distinction as follows, where circles represent core events, and solid horizontal lines between circles represent plan-forming relations. Crucially, there are no such relations in (28b), which is the reason for the absence of the single-event interpretation.¹⁴

(28) a.



b.



At this point, we have developed a recursive model of macroevent formation which accords well with intuitions concerning the delimitations of events, as attested both directly and through Fodor's generalisation. The model has two stages. Firstly, we define a set of *core events*, whose internal structures consist maximally of a process which directly causes a culmination. Secondly, these core events are chained into *extended events* corresponding to an agent's plans. In the following section, this model will be combined with condition (11) to give an account of the cases of legitimate extraction from adjuncts described in section 2.

4. Reaching the parts that syntax cannot reach

Section 2 introduced two major challenges to a syntactocentric theory of locality conditions on extraction from adjuncts. The first concerns the patterns of extraction from bare present participial adjuncts, the acceptability of which is conditioned by the aspectual class of the VP to which the adjunct is attached. The second concerns the general acceptability of extraction of nominal complements from within *in order* clauses. Sharpening the second question a little, we are interested in why it is precisely the class of *in order* clauses that allow such free extraction. I will claim below that the theory of event structure developed in section 3, together with condition (11), automatically predicts such a pattern. Before that, however, I want to see what syntactic alternatives have to say.

In fact, no current purely syntactic theory of locality gets close even to describing the relatively limited set of data given here.¹⁵ Syntactic locality theories can be broadly divided into four

¹⁴Of course, there are still relations of direct causation between adjacent events, which I have represented by a dotted line in (28b). This means that the intermediate process and the final culmination can jointly form a core event, according to the description in (15). This is in conformity with Fodor's generalisation, as a lexical causative description such as *The draught dispersed the smoke* is quite appropriate in such a case.

¹⁵This is not to claim that the data are not well-known, but simply that they have long been seen as a thorn in the side of locality theories. Chomsky (1982), for example, lists examples such as the following, commenting that they 'range in acceptability from fairly high... to virtual gibberish' (Chomsky 1982:72). No satisfying account of this gradience is forthcoming there, however.

(i) a. Here is the influential professor that John went to college in order to impress *e*.

classes, the status of one of which is less clear from a minimalist perspective. Firstly, post-CED theories (Uriagereka 1999; Johnson 2002) work on the assumption that movement is absolutely impossible from certain domains. Secondly, minimality-type theories (Rizzi 1990; Chomsky 1995; Starke 2001) claim that a given movement step is illicit if it crosses a closer element of 'the same type'. On the minimalist reinterpretation of this approach, this reduces to the claim that a feature on a probe can only enter into an agreement relation with that feature on a goal if there is no closer element bearing the same feature. Thirdly, phase-based theories (Chomsky 2000, 2004; Fox & Pesetsky 2005) claim that movement out of certain domains is impossible unless the moved element is close enough to the edge of that domain. Finally, the older, and currently somewhat disfavoured, barriers-type theories (Chomsky 1986; Cinque 1990) suggest that a movement step crossing a given node is illicit unless that node enters into certain relations (e.g. proper government) with other nodes.¹⁶

It should be clear that a post-CED theory is no use to us here. All of the references cited above make extraction from an adjunct impossible, in which case the existence of examples such as (10a) or (17), repeated below, is seriously problematic.

- (10) a. What did John drive Mary crazy [complaining about *e*]?
 (17) a. What are you working so hard [in order to achieve *e*]?
 b. Who did John travel to England [to make a sculpture of *e*]?
 c. Whose attention is John jumping up and down [in order to attract *e*]?
 d. What did you tap your nose [in order to signal to Mary *e*]?
 e. What did you come in [to talk to us about *e* today]?

An alternative post-CED theory would be one in which the criteria for domains allowing extraction are modified in such a way that extraction from an adjunct is uniformly possible. However, in that case, the ungrammaticality of examples such as those in (4) goes unexplained.

- (4) a. *What did you come here [because Mary wanted to talk to you about *e*]?
 b. *What did John go home [after Mary said *e*]?
 c. *What was little Jimmy playing in the sandpit [screaming about *e*]?

What a post-CED theory lacks is a way of stating that extraction from a given domain is sometimes permitted and sometimes impossible, but that seems to be the best description of the facts here. The post-CED approach will not give us a handle on these facts, then.

The minimality-type, phase-based, and barriers-type approaches *can*, at least in principle, capture the fact that extraction from an adjunct is only sometimes legitimate. On a minimality approach, the natural way to capture such a pattern is to associate an \bar{A} feature of the relevant type with the head of all and only adjuncts which do not allow extraction. Perhaps, for a phase-based approach, we could assume that adjuncts are always phases, but that only certain adjuncts have an available edge position, and so only those adjuncts will allow subextraction. Meanwhile,

-
- b. The article that I went to England without reading *e*.
 c. The book that I went to college because I liked *e*.
 d. The man that I went to England without speaking to *e*. (Chomsky 1982:72)

¹⁶Of course, the boundaries between these modes of explanation are not sharp, and many hybrid theories exist. This is a good first approximation to the state of the art, though.

on a barriers-type approach, the opportunities are legion.¹⁷ Without going into details, we may expect height of attachment to affect the possibility of extraction, given that adjunction of a *wh*-phrase to VP, but not IP, is possible in Chomsky (1986). It is also possible to stipulate that some, but not all, adjunct maximal projections allow adjunction, thereby voiding their barrierhood.

However, the minimality-type, phase-based, and barriers-type approaches suffer from a quite complementary problem, already flagged up in section 2. All three of these approaches relate locality effects to properties of the nodes crossed by a given movement step. However, the paradigm in (10) shows that contrasts in grammaticality of extraction from adjuncts cannot always be reduced to properties of the nodes on the path from base position to surface position.

- (10) a. What did John drive Mary crazy [complaining about *e*]?
 b. *What does John do his work [complaining about *e*]?

Faced with this problem, we may wonder whether an analysis in the spirit of the original CED, whereby the distribution of a syntactic relation determines extraction possibilities, could account for this paradigm. In principle, a CED-type analysis is better equipped to deal with apparent locality effects due to properties of the adjunct's sister, as it may be possible to claim that some extra relation (call it the *pseudoargument* relation) holds between an adjunct and (nodes within) its sister in a case like (10a), in addition to any such relations that hold in (10b). However, I suspect that, fully developed, it would amount to little more than a syntacticisation of the semantic story developed here: the distribution of the putatively syntactic pseudoargument relation would in fact be determined on the purely semantic grounds discussed above, leaving this approach with no motivation beyond a desire to keep locality theory in the syntax. In such a situation, I would argue that the burden of proof lies with the more complex syntactocentric theory. However, I have been unable to find any evidence for the syntactic reality of such a pseudoargument relation.

Other approaches also suggest themselves. For example, we can, of course, deny the assumption that the adjunct-external syntax of these two examples is the same. This would allow us to attribute the difference in acceptability between the two examples in (10) to, for example, a syntactic height effect. However, while not impossible, this route once again lacks independent motivation: according to most standard constituency tests, it seems that the two adjuncts in (10) are attached at the same height. Moreover, it should be noted that extraction from relatively high adjuncts is not universally impossible, as many people accept stranding of temporal prepositions, as in (29), which are generally taken to attach in the IP domain. It seems, then, that height alone cannot provide us with all the answers here.

- (29) Which film did John fall asleep [during *e*]?

A further possibility, suggested by an anonymous reviewer, is that we may posit a covert *while* in (10b), which is absent in (10a). We could then link the different extraction patterns to adjunct-*internal* syntactic factors, most naturally the syntactic size of the adjunct. If (10b), containing covert *while*, is a CP, whereas those adjuncts which allow extraction are TPs or smaller, then standard modes of explanation would become available to us. However, this approach, too,

¹⁷See Jones (1987) for a critical survey.

raises many questions. Firstly, we need to explain why, for example, a covert *while* has this effect, but not an equally plausible covert *by* in (10a). Secondly, it would then be unclear why cases of extraction from similar constructions with an overt preposition, as in (7a), are frequently acceptable.

Of course, it is quite possible to stipulate our way through all these difficulties on a syntactic approach. It would appear, though, the end result will be a highly disjunctive theory, with separate conditions applying to different classes of adjunct. What is much harder is to come up with a unified syntactocentric theory with genuine explanatory potential, as more elegant syntactocentric approaches invariably run into serious empirical problems. Now, consider how the event-based approach accounts for this pattern.

The core of the event-based approach lies in condition (11), repeated below.

(11) Events define locality domains for *wh*-movement

A *wh*-chain is legitimate only if the denotation of the minimal constituent containing the head and foot of the chain asserts the existence of a single event.

Fleshing out the notion of *single event*, sections 2 and 3 proposed a two-stage characterisation of events. Firstly, *core events* consist at most of a single process which directly causes a single culmination. Secondly, *extended events* consist of a series of core events corresponding to an agent's plan for reaching the goal, represented as the final core event in the series.

Moreover, we saw in section 2 that the characterisation of core events alone was sufficient to derive the pattern in (10). (10a) is grammatical as a result of the adjunct event's coming to be identified as the preparatory process leading to the culmination specified in the matrix VP. However, no such identification is available in the cases of (10b) and (16), and as a consequence, (11) cannot be satisfied in those cases.

The extension to extraction from *in order* clauses should now be clear. *In order* clauses describe an agent's plans, and consequently, the matrix and adjunct events together form an extended event. This is sufficient to permit extraction from the adjunct, whatever the aspectual class of the event-denoting constituents in question.¹⁸ Whereas extraction from bare present participial adjuncts, which can only form core events with their sisters, is accordingly contingent on various factors concerning the rather limited internal structural possibilities of core events, extraction from *in order* clauses can take advantage of the fact that the extended event formation

¹⁸Of course, certain combinations of events in *in order* constructions will be infelicitous, regardless of *wh*-movement, because the combination of the two events in question do not correspond to our real-world knowledge of what constitutes a reasonable plan. Crucially, though, even the declarative versions of such sentences, as in (ia), are infelicitous. Extracting out of the adjunct, as in (ib), doesn't make things any worse.

- (i) a. #John is smiling in order to fry an egg.
 b. #Which egg is John smiling in order to fry?

I am aware of one exception to the claim that extraction of DPs from *in order* clauses is unproblematic. This is based on uses of these clauses, discussed in Culicover & Jackendoff (2001), in areas such as literary criticism to describe an author's intention at a given point in the text. Such examples do not require agentivity, or even animacy, on the part of the subject, and the adjunct describes the *author's* plan, not the subject's. In such cases, extraction from the *in order* clause is impossible, as shown in (ii).

- (ii) a. The ship sinks [in order to further the plot].
 b. *What does the ship sink [in order to achieve *e*]?

At present, I have no satisfactory explanation for this fact.

process can chain together any class of core events, which guarantees that the matrix and adjunct events will jointly form an extended event in this case. Unlike bare present participial adjuncts, then, satisfaction of (11) by, and consequently extraction from, *in order* clauses is predicted to be automatically possible.

There remains one question, which is why bare present participial adjuncts and *in order* clauses should align themselves with core events and extended events in this way. I claim that this is due to the overt forms that the two classes of adjuncts take. Bare present participial adjuncts are distinguished by not having any overt marker of the precise semantic relation between the matrix and adjunct events. In contrast, *in order* clauses are distinguished by the presence of an initial *in order to* (or just *to*), which clearly signals that the matrix and adjunct events are related as the initial event and the goal of a plan.

The representation of *in order* clauses as extended events is unsurprising, in that case, but why should bare present participial adjunct constructions, without any overt marking of the relation between matrix and adjunct events, not be interpretable as extended events? I believe that this is because of a basic fact about plans: they sometimes fail. In other words, there is no guarantee from the use of an *in order* clause that the goal event occurred just because the initial event did, but the two events in a bare present participial adjunct construction are not separable in this way. So while (30a) is a quite acceptable utterance, (30b) is contradictory.

- (30) a. John came to England in order to meet the Queen, but he never got to meet her.
 b. #John drove Mary crazy complaining about the weather, but he never {complained about the weather / drove Mary crazy}.

In this case, the restriction of bare present participial adjuncts to core events makes sense if we adopt the following natural hypothesis.¹⁹

- (31) Modality must be overtly signalled.

(30) shows that there is a modal element to the relation signalled by *in order* which is absent from the relation among events in a bare present participial adjunct construction. If the matrix event in a bare present participial adjunct construction occurred, then so did the adjunct event, but this does not hold for the *in order* case. This is as expected if (31) holds.

Now, let's consider the relations among subevents that hold in core events, and in extended events. We have seen that subevents of extended events are related by the forward- and backward-chaining relations that govern plan formation. We have also seen from (30a) that such relations are modal in nature. However, the only relation that holds among events in core events is direct

¹⁹A reviewer asks whether we may expect that modality will facilitate extraction in general. Although little rests on the answer for the issues narrowly at hand here, it seems in fact that, in at least one case, there is a connection between modality and extraction. This is the factive island phenomenon discussed by Erteschik-Shir (1973), and illustrated below.

- (i) a. Who did John say [that Mary kissed *e*]?
 b. *Who did John regret [that Mary kissed *e*]?

Although space precludes a discussion here, an extension of the event-structural condition is proposed in Truswell (2007b:§3.3) which makes crucial use of the relativisation of the kissing event to John's belief worlds in (ia), but not in (ib), in the same way that the extended events hypothesised to lie behind extraction out of *in order* clauses are relativised to John's planning worlds, however they may ultimately be formalised.

causation. This is not modal in nature. A basic definition of causation (see, for example, Lewis 1973 for a more subtle approach along these lines) is as follows.

(32) An event e_1 *causes* an event e_2 iff:

- a. e_1 and e_2 both occur;
- b. If e_1 hadn't occurred, then, all else being equal, e_2 wouldn't have occurred either.

This is exactly the sort of inseparability that we observed with respect to the two subevents of a bare present participial adjunct construction in (30b). If it is legitimate to adopt (31), then, it follows that the subevents in a bare present participial adjunct construction cannot jointly form an extended event, but they *can* form a core event.²⁰

Summing up, this section has shown that any orthodox syntactic locality theory encounters serious problems in accounting for even the few pieces of data presented here. However, the event-based approach to locality, supplemented by one natural assumption, (31), about overt signalling of modal elements, breezes through the data.

5. Conclusion

The empirical impetus for this paper comes from a few clear exceptions to the otherwise well-motivated generalisation that adjuncts are strong islands. The interest of these exceptions is that, rather than necessitating some small amount of tinkering with the theory, they have led to a radically different view of the ill-formedness of many extractions from adjuncts. Section 4 showed that the cases where extraction is possible form natural classes on an independently motivated view of the internal structure of events, but not on any current view of phrase structure.

Where does this leave the overall architecture of a theory of locality? Certainly, adjuncts are not (always) strong islands — the counterexamples scattered throughout this paper are sufficient to dispel that claim. However, we also cannot expect to remove all of locality theory from syntax. This is clearly shown by the effects noted by Cinque and discussed in section 2, the upshot of which is that even when they allow subextraction, adjuncts still show many typical properties of *weak* islands. So, for example, we saw in (7), repeated below, that extraction of anything other than a DP from an adjunct is impossible, as is typical of weak islands.

- (7) a. The man that I went to England without speaking to e
 b. *The man to whom I went to England without speaking e (Chomsky 1982:72-3)

Such sensitivity to syntactic category would be hard to reproduce in a purely semantic theory. It seems, then, that classic locality effects are partly a matter of syntax, and partly a matter of

²⁰This does not mean, of course, that they *have to* form a core event. In a case like (i), the relation between the matrix and adjunct events is in no way causal. Instead, it is more accurately characterised as conjunction, with additional appropriate restrictions on spatiotemporal location of the two events.

(i) John works building igloos.

However, in such a case, the two events do not jointly form a single macroevent. Accordingly, (11) is not satisfied and extraction is impossible.

(ii) *What does John work building?

semantics. As ever, there are two obvious ways to accommodate such a result within the theory. Either we can give up on the standard Y-model, in favour of a parallel architecture in which the logical priority of syntax is reduced or abolished (see, for example, Jackendoff 2002), or we can conceive of the semantic component as a filter, operating on an overgenerating syntactic component by rejecting those sentences which do not comply with condition (11). Choosing between two such alternatives is far beyond the scope of this paper, however.

Acknowledgements

Thanks to Ad Neeleman for putting up with the development of these ideas over the last couple of years. Thanks also to David Adger, Richard Breheny, Annabel Cormack, Heidi Harley, Kyle Johnson, Hans van de Koot, Øystein Nilsen, Gillian Ramchand, Mark Steedman, Hiroyuki Uchida, Phil Wolff, an anonymous reviewer, and audiences at Newcastle, Queen Mary, and Utrecht for comments and suggestions. I am also grateful to Phil Wolff and to Elsevier for permission to reproduce the illustrations contained in this paper. This work was initially supported by an AHRC studentship and is now supported by a Wingate scholarship. And last but not least, thanks to all at ConSOLE for making it such an enjoyable weekend.

Robert Truswell

University College London

r.truswell@ucl.ac.uk

<http://www.phon.ucl.ac.uk/home/robert/>

References

- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy* 9:1, pp. 5–16.
- Bittner, M. (1999). Concealed causatives. *Natural Language Semantics* 7:1, pp. 1–78.
- Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. MIT Press, Cambridge, MA.
- Chomsky, N. (1986). *Barriers*. MIT Press, Cambridge, MA.
- Chomsky, N. (1995). *The minimalist program*. MIT Press, Cambridge, MA.
- Chomsky, N. (2000). Minimalist inquiries: the framework. Martin, R., D. Michaels & J. Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, MIT Press, Cambridge, MA, pp. 89–115.
- Chomsky, N. (2004). On phases. Ms., Massachusetts Institute of Technology.
- Cinque, G. (1990). *Types of \bar{A} -dependencies*. MIT Press, Cambridge, MA.
- Culicover, P. & R. Jackendoff (2001). Control is not movement. *Linguistic Inquiry* 32:3, pp. 493–512.
- Dowty, D. (1979). *Word meaning and Montague Grammar: The semantics of verbs and times in generative semantics and Montague's PTQ*. Reidel, Dordrecht.
- Erteschik-Shir, N. (1973). *On the Nature of Island Constraints*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Fodor, J. A. (1970). Three reasons for not deriving “kill” from “cause to die”. *Linguistic Inquiry* 1, pp. 429–438.
- Fox, D. & D. Pesetsky (2005). Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31:1, pp. 1–45.
- Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9, pp. 183–221.
- Higginbotham, J. (2000). Accomplishments. *Proceedings of GLOW in Asia II*, pp. 72–82.
- Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

- Jackendoff, R. (2002). *Foundations of Language*. Oxford University Press, Oxford.
- Johnson, K. (2002). Towards an etiology of adjunct islands. Ms., University of Massachusetts, Amherst, MA, available at <http://people.umass.edu/kbj/homepage/Content/Etiology.pdf>.
- Jones, C. (1987). P for proper governor. *Proceedings of the 6th West Coast Conference on Formal Linguistics*, pp. 115–129.
- Jones, C. (1991). *Purpose clauses: Syntax, thematics, and semantics of English purpose constructions*. Kluwer, Dordrecht.
- van de Koot, H. (2004). Strong islands and the nature of dependencies. Ms., University College London, available at http://www.phon.ucl.ac.uk/home/hans/strong_islands.pdf.
- Levine, R. & I. Sag (2003). Some empirical issues in the grammar of extraction. Müller, S. (ed.), *Proceedings of the 10th international conference on Head-driven Phrase Structure Grammar*, CSLI, Stanford, CA, pp. 236–256.
- Lewis, D. (1973). Causation. *Journal of Philosophy* 70:17, pp. 556–567.
- Link, G. (1983). The logical analysis of plural and mass terms: A lattice-theoretical approach. Bäuerle, R., C. Schwarze & A. von Stechow (eds.), *Meaning, use and the interpretation of language*, Walter de Gruyter, Berlin, pp. 303–323.
- Parsons, T. (1990). *Events in the semantics of English: A study in subatomic semantics*. MIT Press, Cambridge, MA.
- Pustejovsky, J. (1991). The syntax of event structure. *Cognition* 41:1, pp. 47–81.
- Rizzi, L. (1990). *Relativized Minimality*. MIT Press, Cambridge, MA.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Starke, M. (2001). *Move dissolves into Merge: A theory of locality*. Ph.D. thesis, University of Geneva. Available from <http://ling.auf.net/lingBuzz/000002>.
- Steedman, M. (2002). Plans, affordances, and combinatory grammar. *Linguistics and Philosophy* 25:5/6, pp. 725–753.
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science* 12:1, pp. 49–100.
- Truswell, R. (2007a). Extraction from adjuncts and the structure of events. *Lingua* 117:8, pp. 1355–1377.
- Truswell, R. (2007b). *Locality of wh-movement and the individuation of events*. Ph.D. thesis, University College London. Available from <http://www.phon.ucl.ac.uk/home/robert/thesis.pdf>.
- Uriagereka, J. (1999). Multiple spell-out. Epstein, S. & N. Hornstein (eds.), *Working minimalism*, MIT Press, Cambridge, MA, pp. 251–282.
- Vendler, Z. (1957). Verbs and times. *Philosophical review* 66:2, pp. 143–160.
- Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition* 88:1, pp. 1–48.
- Zwart, J.-W. (2007). Layered derivations. Paper presented at UCL, handout available at <http://www.let.rug.nl/zwart/docs/ho07actl.pdf>.