Chapter 3
Modals, Small Clauses and Causatives

Confusion in her eyes that says it all—she’s lost control.
— Joy Division, “She’s Lost Control”, Unknown Pleasures

3.1 Introduction
In this chapter, several other constructions in English involving “obligatory control” will be discussed and given an analysis consistent with the analysis given in chapter 2. In that chapter, we saw that by allowing multiple theta role assignments to the same Case-chain, we can dispense with a separate theory of obligatory control. In this chapter, we will consider constructions—including Modal verbs, small clauses, and causatives—in which head movement, Case-movement and theta role assignment interact.

This chapter is organized as follows: in §3.2 we will consider the relation of head-movement and argument, and this will lead to an investigation of the interaction of head-movement with theta role assignment. We will also solve a remaining problem from chapter 2. In §3.3, we will extend this analysis to small clauses, in §3.4, to modal verbs, and in §3.5 to causative constructions in English (and in a few other languages). In §3.6 we review this chapter’s results.

3.2 Head Movement and Theta Role Assignment
In chapter 2, we argued that theta role assigners can assign their theta roles to any Case-chain containing a position within the assigner’s minimal domain. Although head movement is recognized in the definition of minimal domain, we did not consider the possibility that head movement might otherwise affect theta-role assignment.

In this section, we will propose that feature checking and theta-role assignment are
closely linked. In fact, this is only a small extension to the so-called “Visibility Requirement” that a DP needs to be Case-checked in order to be assigned a theta role at LF. In chapter 2, the only type of feature checking we considered was Case-checking movement of a DP to the Specifier of an Agr phrase. This is a special case of the more general principle that all argument phrases must check some feature—a “licensing” feature—in order to be available at LF to be assigned any sort of theta role—this is, to have that phrase’s interpretation composed with the interpretation of the rest of the sentence. This implies that an argument phrase must be in a checking position at LF to be assigned a theta role—although not necessarily the checking domain of the theta-role assigner. This movement to check a licensing feature may be overt (before Spell-Out, e.g. the movement of a subject DP to SpecAgrSP) or covert (after Spell-Out). As we will see in this chapter, examples of licensing movement include:

- movement of the head of a small clause, head-adjoining covertly with the verb which governs the small clause (which we will see more of below)—that is, “covert incorporation”;
- the covert head-adjunction of “main verbs” with modals; and
- the head-movement of a “caused” verb to adjoin to the causative verb or morpheme.

This is a natural extension of the movement operations we have already examined in chapters 1 and 2, and which are common in Minimalist analyses. We have already seen covert movement of the verb to AgrO˚ in chapter 2; this post-Spell-Out head movement allows the verb to assign a theta role to a DP in SpecAgrOP.

Up to now, Case-checking is the only feature-checking operation we have seen. However, there are many other constituents beside DP’s which receive theta role assignments—CP/AgrSP/TP, AP, and PP, for example—and no not have Case features needing to be checked. However, all of these non-Case-checked phrases have one thing in common: they are selected to be the complement of a verb. Therefore, to be assigned a
theta role, a constituent must either check its Case features or check other kinds of selectional features. Certain types of these selectional features have already been discussed: verbs selecting for different features (i.e. types) of TP clauses, for example, and we will see many more examples in the remainder of this thesis. Sometimes (as for Case-movement) there are specific sites—the Specifiers of certain projections—at which Case features can be checked. However, for some selectional features there will be fewer choices about how to check them. Primarily, many selectional features are checked by covert head movement of the head of the selected phrase to the head which selected that phrase. We can stipulate initially that only certain feature are licensing features.

(1) Licensing Features

Licensing Features, which make a constituent Visible for theta assignment at LF, include the following features and feature-checking operations:

- Case and Agreement, and/or a D-feature (that is, the EPP-satisfying D-feature of Chomsky 1995) for DP’s;
- Tense for TP’s;
- Agreement for AP’s; and
- Selectional features for a selected maximal projection.

It is perhaps not surprising to assert that the heads of small clauses, modal verb complements and causative clauses adjoin to their governing verb to check a feature. Certainly there are many languages which have overt incorporation of a verb into a causative morpheme, (see Baker 1988 and references therein) which makes it natural to propose that the remaining languages have covert causative incorporation.

Likewise, most of the West Germanic languages (with the notable exception of English) seem to have overt movement of an infinitive verb to the governing modal verb. The literature on West Germanic verbal raising is extensive and remains an active area of
research; for example, see Evers (1975), Bresnan et al. (1982), den Besten and Edmondson (1983), Haegeman and van Riemsdijk (1986), Johnson (1988), Kroch and Santorini (1991), Zwart (1993) and Vanden Wyngaerd (1996). In Dutch and German and their dialects, there are a variety of dialect-specific sets of rules for determining the properties—the structure and the order of the verbs—of the clause-final verb cluster. In German, consider the following sentence:

(2) Ich glaube nicht, daß er die Lieder wird haben singen können

I do not believe that he will have been able to sing the songs.

The embedded clause-final verb cluster *wird haben singen können* is required to be ordered in just this way, which is compatible with neither a head-final specification (which would predict the sentence’s embedded clause to be *daß er die Lieder singen können haben wird*) or a head-initial one (which would predict the sentence’s embedded clause to be *daß er wird haben können singen die Lieder*). Therefore, it is assumed that overt movement of all the verbs into a single cluster is the only way to allow this variety of almost morphological constraints to operate. If such head-movement is overt in West Germanic, therefore one may suppose that other languages must do such movement covertly.

What is the effect of this proposed head-movement, though? Up to now, it would have no effect on theta-role assignment, but now we will propose a theory of the effect of head movement on theta-role assignment. Informally, it states that two theta-role assigning heads must not be structurally too close together; unless they are sufficiently independent, both of them cannot assign a theta role to the same object. Let us define the following principles.
(3) **Theta-Role Assignment by Adjoined Heads**

If two theta-role assigning heads $\alpha$ and $\beta$ are amalgamated at LF, then $\alpha$ and $\beta$ cannot both assign a theta role to the same LF object.

(4) **Amalgamation**

$\alpha$ and $\beta$ are amalgamated iff:

a. $\alpha$ adjoins to $\beta$, or

b. $\alpha$ is contained immediately within $\gamma$, $\gamma$ adjoined to $\beta$, and $\gamma$ has a null interpretation at LF.

With these definitions, we prevent heads which are “too close” to each other from assigning theta roles to the same LF object. This is only a minor extension of Thematic Uniqueness, defined in chapter 2 ex.7 and repeated here:

(5) **Thematic Uniqueness**

For a theta-role assigning head $\eta$ and a legitimate LF object $\alpha$, no more than one of the theta roles of $\eta$ may be assigned to $\alpha$.

Amalgamation extends this to the situation where two theta-role assigning heads are adjoined to each other — or virtually adjoined to each other from the point of view of the LF interpretation apparatus.¹

¹As discussed in a footnote in chapter 2, the deep motivation for Amalgamation is the empirical observation that there is a complementary distribution between a verb selecting a LF-interpretable head and a verb having checkable Case-features. In reality, though, this complementarity should not be surprising. After all, when a noun adjoins to a verb in Noun Incorporation, the verb loses its ability to check Case, because those Case-features are used to check the Case features of the incorporating N. Perhaps we can broadly assume that LF-interpreted heads need to check Case when they adjoin to the verb which selects them, in accordance with the Visibility Condition. If true, it is then clear that heads without an LF-interpretation — by definition invisible at LF — do not need to check Case, and therefore must not check Case. If we stipulate that verb with Case-features must check their Case-features in order to be interpretable at LF, then we derive the correct properties of both raising and control verbs. Thus, Amalgamation may be considered not an completely new and unmotivated principle of grammar, but a direct consequence of the Visibility Condition.
Motivating these definitions will take up the rest of this chapter, but first, we can now clear up some unfinished business from chapter 2, which is a particular example of these definitions at work.\footnote{An alternative approach to the problems to be discussed in this chapter is to assume that a raising verb obligatorily has Case-features which need to be checked by LF for the derivation not to crash. However, such an requirement is unusual from the standpoint of the Minimalist program, as well as being unnecessary in other constructions (and even making incorrect predictions, as in example (14) in this chapter). So we will leave it aside here.}

As we discussed at the end of §2.3, the sentence *Jamie, believes t, to have left is predicted to be grammatical given the framework we argued for in that chapter, for the same reason that the sentence Jamie, wants t, to have left is grammatical—in the framework of chapter 2, there is no way to prevent a ECM verb from also optionally being a control verb. However, the sentences are not exactly parallel. As Stowell (1982) argued, although the tense operator of control infinitives is not specified for [± Past], it does specify that the time frame of the infinitival clause is unrealized with respect to the tense of the matrix in which it appears. Alternately, the tense of a to-infinitive is a “possible future.” Consider the following sentences (from Stowell 1982:563 ex. 8 & 9, changing the controlled PRO to a trace):

(6)  
\begin{itemize}
  \item a. Jenny\textsubscript{i} remembered [t\textsubscript{i} to bring the wine] 
  \item b. Jenny\textsubscript{i} remembered [t\textsubscript{i} bringing the wine] 
\end{itemize}

(7)  
\begin{itemize}
  \item a. Jim\textsubscript{i} tried [t\textsubscript{i} to lock the door] 
  \item b. Jim\textsubscript{i} tried [t\textsubscript{i} locking the door] 
\end{itemize}

In each pair of sentences, the to-infinitive is understood to be unrealized with respect to matrix’s tense: thus in (6a) Jenny has not brought the wine when she remembered to do so, and in (7a) Jim need not actually succeed at locking the door. Compare these interpretations with those of (6b) and (7b), where as Stowell notes, “the understood tense of the gerund is completely malleable to the semantics of the governing verb.”
shows that *to*-infinitives have a regular internally specified “unrealized” tense. In contrast, Stowell notes that raising verbs do not have a regular internally specified “unrealized” tense; instead, the understood tense of these complements with respect to the tense of the matrix is determined largely by the meaning of the matrix verb. Consider the following ECM sentences (from Stowell 1982: 565-566 ex. 12a & 13a,b).

(8)  a. John considers [himself to be the smartest]
    b. I expect [John to win the race]
    c. I remember [John to be the smartest]

In (8a) the infinitive is understood to have a present tense, in (8b) the infinitive is understood to have a future tense, and in (8c) the infinitive is understood to have a past tense. There is no important relevant structural or lexical difference between these infinitive clauses, so the only reasonable explanation for these differences is that the matrix verb is supplying the tense interpretation.

Let us construe these results in the following way with respect to the mechanisms allowed by the Minimalist Program. Let us assume that every embedded T˚ is required to move to a matrix T˚ to check its tense features, and thereby receive an interpretation at LF. This is required by the need for arguments become licensed by moving to check selectional features and thereby becoming Visible for theta-role assignment at LF, which was proposed earlier in this chapter. However, in certain cases (e.g. raising verbs), the matrix verb selects an embedded T˚ which does not in itself have an interpretation at LF, although it has the same syntactic requirements as other embedded T˚ heads. In these cases, the theta-role assignment possibilities for the adjoining heads might be restricted by the definition of amalgamation in (3), which prevents two heads assigning a theta role to the same LF object if the two heads are separated only by a head which has the null interpretation at LF. In light of this, consider the following LF structure for the sentence
similar to one we considered in chapter 2, Andy wants to be here, this time with a somewhat more complete display of the head movement involved.

(9)

It is critical to note in (9) that \(t_{be}\) adjoins to \(t_{tense}\), and \(t_{tense}\) adjoins to \(t_{wants}\). If we stipulate \(wants\) selects a complement clause with an uninterpreted tense specification (which is motivated by its observed tense semantics), that \(T^\circ\) is to be interpreted at LF. Therefore, \(be\) and \(wants\) are not amalgamated as defined in (4), and both \(be\) and \(wants\) can assign theta roles to the Case-chain \([\{DP \text{Andy}\}, t'' \ t', t]\) in (9). Thus for this sentence, we derive the same result we got in chapter 2. However, consider the LF of the sentence *Andy
believes to be here, which was incorrectly predicted to be grammatical by the framework of chapter 2

(10)

We see that (10) has the same structure as (9), modulo lexical insertion. However, the verb believes selects an embedded T˚ with a null interpretation since believe supplies the tense interpretation of the embedded clause. Therefore, after the adjunction of be to T˚, and T˚ to believes, we get the head adjunction structure \( [\_v \text{ believes} [\_T T˚ [\_v \text{ be}]]] \), and since this T˚ has a null interpretation at LF, by (4) believes and be are amalgamated and cannot both assign a theta role to the same Case-chain. Since there is only one Case-chain
in (10) to which both verbs assign a theta role, (10) must be ungrammatical.

This is a desirable result, since it would be a significant problem if our theory allowed all ECM verbs also to be subject control verbs. However, as we will see in the rest of this chapter, there are several other applications of these principles which considerably extend the empirical range of the theory presented here. First we consider Small Clauses in English, and why they do not allow control.

3.3 Small Clauses

Small clause constructions in English have a number of interesting properties. Consider the following set of data:

(11)  a. Andy considered [IP Jamie to be intelligent]
      b. Andy wanted [IP Jamie to be intelligent]
      c. *Andy₁ considered [IP t₁ to be intelligent]
      d. Andy₁ wanted [IP t₁ to be intelligent]

(12)  a. Andy considered [SC Jamie intelligent]
      b. *Andy wanted [SC Jamie intelligent]
      c. *Andy considered [SC t intelligent]
      d. *Andy wanted [SC t intelligent]

The pattern of grammaticality in (11) is familiar from chapter 2. The matrix verb in (11a&c), consider, is a raising verb—that is, it selects for a to-infinitive with an interpretatively transparent T°. In (10a), the DP Jamie is assigned the sole theta role of the embedded predicate be intelligent, and the DP Andy is assigned the agent theta role of consider. However, (11c) is structurally parallel to (10); the only available DP chain headed by Andy cannot be assigned theta roles by both considered and be intelligent. This is because the head adjunction structure they are in at LF is as follows:
We know that the T° head has the null interpretation at LF, because considered is a raising verb and because of its tense semantics, which we already discussed. Therefore, in the head-adjunction structure in (13) the two verbs considered and be intelligent are amalgamated and cannot both assign a theta role to the same Case-chain. Since there is only a single Case-chain available for theta-role assignment, (11c) cannot be grammatical.

The sentences (11b) and (11d) are completely analogous to sentences we discussed in chapter 2. In particular, since the matrix verb wants selects a to-infinitive with an “unrealized” tense which has an LF interpretation, the two theta-role assigning verbs wants and be intelligent are never amalgamated and hence may assign their theta roles to the same Case-chain, as in (11d).

With all this in mind, let’s consider the structure of small clauses. Let us assume that all small clauses are XP’s—that is, a small clause like \([\text{SC} \text{ Jamie intelligent}]\) in (12a) actually is an AP with the structure \([\text{AP} \text{ Jamie} \ [\text{A} \text{ intelligent}]\]) where intelligent is the head of the AP and Jamie is in the AP’s specifier position. The head of the AP, intelligent, must move to check its Agreement features against the features of the main verb, as well as to check its selectional features and to be Visible for theta-role assignment at LF. This will be a head-adjunction movement, not XP movement to a specifier position, since no such positions are available for the kind of features being checked by the small clause head. Of course, this small clause head movement is covert in English.\(^3\) If this is true, and there are no other intervening LF-interpreted heads, then the head of the small clause and the main verb will be Amalgamated, and therefore both the main verb and the head of the

\(^3\) Although it seems to be semantically real, as well as being required by the syntax. Specifically, the usual analysis of small clauses in Montague semantics requires that the small clause head and the main verb are semantically composed first, then that composed semantic function is applied to the subject of the small clause. This order of function-applying operations may be directly derived from the feature-checking structures proposed here.
small clause cannot assign theta-roles to the same Case-chain. This is exactly the same reason that made control impossible with verbs like *believe*—and therefore is directly related to the difference between raising and control constructions. Therefore the theory predicts that control is impossible in small clause constructions.

This prediction is in fact correct. Recall the small clause data given above. In (12) the small clauses are AP’s, and the DP *Andy* starts in SpecAP in each sentence in (12). In each of the sentences in (12), then, the adjective *intelligent* must adjoin to the matrix verb—*considered* in (12a&c) and *wanted* in (12b&d)—in order to be licensed and hence Visible for theta assignment at LF. Thus, (12a) is a typical small clause, with *considered* assigning its theta role to the Case-chain headed by *Andy* and *intelligent* assigning its theta role to the DP *Jamie*; the same may be said for (12b). However, (12c&d) are both ungrammatical, and for precisely the same reason. Consider the following LF structure for (12c).

\[
\text{---}
\]

\text{The single question mark beside (12b) means that it is slightly odd, at least for the author. However, it does seem to be grammatical in most circumstances. Its slight deviance is likely to be entirely unrelated to the issues at hand; in this example, the change-of-state semantics which are induced by *wanted* may force the semantics of the small clause and of *intelligent* to be coerced from a purely stative one.}
In (14) we see that *intelligent* adjoins directly with *considers*, since it is the head of the small clause and the small clause must be licensed at LF in that way. However, this means that *intelligent* and *considers* are amalgamated by the definition in (4) and therefore they cannot both assign theta roles to the Case-chain \([\text{DP } \text{Andy}], t\prime, t\); that is ruled out by (3). Therefore, there is no correct derivation at LF for (12c). The same reasoning holds for (12d), since the LF structure of (12d) is identical to (14), the LF structure of (12c), modulo lexical insertion of the matrix verb. Here, the differences between a raising verb and a control verb do not matter, since they are different only in their selection of the properties of embedded T’, which is not applicable when their complement is a small
clause and has no TP.

Similarly, we predict that raising-to-subject verbs would allow small clauses, because the matrix verb does not assign a theta role to the raised DP; therefore, Amalgamation cannot make a difference. This is in fact the case, as indicated by (15).

(15) Jamie₁ seems \([_{sc} t_1} \text{ intelligent}\)

Although intelligent must be adjoined to seems at LF in order to license the small clause (and hence the two heads are amalgamated), it does not interfere with theta role assignment since seems does not assign a theta role to the raised DP’s Case-chain; only intelligent does. Therefore the Amalgamation principle in (3) is trivially satisfied, and this prediction is borne out. Moreover, we also expect that object control verbs would never allow small clauses, as in (16).

(16) *Andy persuaded Mary₁ \([_{sc} t_1} \text{ intelligent}\)

This prediction is also true. In (16), the head of the small clause intelligent must adjoin to the matrix verb persuaded at LF for the small clause to be licensed. Without an intervening head, they become amalgamated and therefore cannot assign theta roles to the same Case-chain headed by \([_{dp} Mary]\). Therefore (16) is not grammatical.

So, we have answered the question posed at the beginning of this section: control is impossible out of small clauses because the head movement required by the head of the small clause is incompatible with the theta role assignment requirements of the control verb. Next we consider a different class of constructions generating superficially similar sentences which nonetheless have very different behavior: modal verbs.
3.4 Modals and Head Movement

Modal verbs have a number of properties which distinguish them from other kinds of verbs. In this section, we will concentrate on English modal verbs, but many of the characteristic properties of modal verbs largely hold cross-linguistically. Details of the cross-linguistic properties of modals, especially in the Scandinavian languages, can be found in Thráinsson and Vikner 1995.

Several properties characterize modal verbs and differentiate them from “regular” verbs. English modals are followed by non-finite forms of the verb without the presence of the non-finite marker *to*:

\[(17) \quad \begin{align*}
\text{a. } & \text{Andy can leave} \\
\text{b. } & \text{*Andy can to leave} \\
\text{c. } & \text{*Andy wants leave} \\
\text{d. } & \text{Andy wants to leave}
\end{align*}\]

In the Scandinavian languages, the situation is more complex. Some of the Scandinavian modal verbs take bare infinitival complements, while others do not. Which ones do varies from language to language.

\[(18) \quad \begin{align*}
\text{a. } & \text{Jeg vil (*at) gå hjem (Danish)} \\
\text{b. } & \text{Ég vil (*að) fara heim (Icelandic)} \\
& \text{I will to go home} \\
& \text{‘I want to go home.’} \\
\text{c. } & \text{Dette kan (*at) gå galt (Danish)} \\
\text{d. } & \text{Þetta kann *(að) fara illa (Icelandic)} \\
& \text{this can to go badly} \\
& \text{‘This may go wrong.’}
\end{align*}\]
Modals in some languages have differences in their agreement properties compared to regular verbs (e.g. note the missing “-s” ending on the English modals above). Modal verbs may also be restricted in where they can appear compared to other verbs, and they have a number of the functions of auxiliary verbs; for instance, with respect to VP ellipsis and tag questions—in English, modals do not require “do-support” for these constructions. Also, in most dialects of English, it is not possible for a modal verb to be immediately embedded under another modal. For example:

(19) *John will can eat lunch soon.

There are dialects of English, as well as many if not all Scandinavian languages (see Vikner 1988 and Thráinsson and Vikner 1995) where double modals are acceptable. Although we will not directly consider them here, there are significant restrictions on the appearance of these double modals which seem closely related to the constraints relating to theta-role assignment and head movement seen in this chapter. Thus, data concerning double modals may be reconciled to the analyses of this thesis.

In addition, it is well known that there are two distinguishable classes of modal verbs, known as “root” and “epistemic” modal verbs.\(^5\) Broadly, epistemic modals modify the truth value of the clause containing the modal; on the other hand, root modals express some sort of obligation, necessity, possibility, volition, ability, or permission, usually, but not always on behalf of the subject of the clause.\(^6\) Therefore, in a sense epistemic modals act analogously to raising verbs, in that they do not theta mark the subject of the sentence. By the same token, however, root modals can be said to assign a theta role to the subject, in addition to the theta role assigned to the subject by the main verb (which is argued for

\(^5\)In fact, it is the possibility of root and epistemic meaning which Thráinsson and Vikner (1995) argue defines modal verbs.

\(^6\)Consider, for example, the sentence “This book must be read immediately.” The unexpressed agent of the passive verb be read seems also to be the locus of the obligation.
by Thráinsson and Vikner 1995, among others). Under our analysis of control, root modals are analogous to control verbs with respect to theta-role assignment.

In the theory presented here, it will be an elementary matter to state the differences and the similarities between root and epistemic modal verbs. It will be argued that root and epistemic modals are structurally identical; the only difference coming from the assignment of an extra theta role by a root modal (to the subject). This is an significant advantage of this theory, and it is difficult to express the same insight in a standard GB-Theory or Minimalist framework assuming some version of Control Theory because the requirement that PRO be ungoverned forces raising and control predicates to have different structures. It follows that root and epistemic modals have very different structures, which is implausible—although it may be true. In the remainder of this section, we show how this analysis explains the properties of modals and we will argue for the correctness of this analysis.

To begin, it is necessary to determine the structure of the complements of modal verbs—precisely, what functional categories are in the modal’s complement clause. For English, it is clear that the verb forms which follow a modal are the same as the forms following to in a non-finite context. We will assume that modals are similar to other non-Case-checking verbs. However, instead of selecting an IP, TP or CP complement like other verbs, they select a “bare” VP complement. (Following Thráinsson 1996, an AgrOP projection may be inserted between the modal and the VP if and only if it is necessary, as was discussed previously.) For example, consider the sentence Andy can see Jamie, with the following LF structure:
There are two Case-chains in this LF structure: one is \{[DP Andy], t’, t\}. It is attracted by the strong Case features of T’ (after T’ is adjoined to AgrS’) and [DP Andy] moves to the Case-checking position SpecAgrSP before Spell-Out. The other Case-chain in (20) is \{[DP Jamie], t\}; it is attracted by the weak Case features of the verb see (in the V’+AgrO’ head-adjunction complex) and therefore [DP Jamie] moves to SpecAgrOP to check its Case features after Spell-Out. The Case-chain \{[DP Andy], t’, t\} is in the Theta Domain of both see and the modal verb, and the Case-chain \{[DP Jamie], t\} is only in the Theta Domain of the verb see. Therefore, since \{[DP Andy], t’, t\} is Superordinate to \{[DP
Jamie], t}, the former Case-chain is assigned the agent theta role of *see* and the latter is assigned its patient theta role.

It is interesting to note, concerning this analysis, that the modal verb *can* is ambiguous between its root and epistemic readings. In its epistemic reading, the sentence means roughly, ‘it is possible for Andy to see Jamie,’ while in its root reading it means, ‘Andy is able to see Jamie.’ That is, the former denotes possibility while the latter denotes ability. In the structure (20), the subject Case-chain \{[DP Andy], t', t\} is in the Theta Domain of the modal verb, and hence it is possible for the modal to assign a theta role to that Case-chain. However, theta-role assignment is not obligatory if there is no theta role for the modal verb to assign. If *can* is chosen from the lexicon as a epistemic modal, then no theta role will be assigned, and if *can* is a root modal then it will assign its theta role in the only possible way—to the Case-chain \{[DP Andy], t'', t\}.

For the derivation in (20) to work, however, there is also head movement to consider. Before Spell-Out, the modal *can* adjoins to the T° head. Following that, the complex [T° + modal] head adjoins to AgrS° head. After Spell-Out, the verb *see* adjoins to AgrO° and the [AgrO° + *see*] head complex adjoins to the [AgrS Modal [AgrS T AgrS]] head complex. In this way, all the heads can check their features and become Visible for theta assignment at LF, as well as checking their selectional features. The head-adjunction structures proposed here for analysis of modal constructions—especially the covert head-adjunction proposals—are quite abstract and apparently ad-hoc. It has been recognized for a long time that types of movement that are covert in one language may be overt in another. So, we should be able to find a language in which the head-movement we see in (20) is actually done before Spell-Out and thus is directly observable.

As a matter of fact, we do. It was mentioned in §3.2 that West Germanic languages other than English have overt movement of a verb to a governing modal verb. In embedded

---

7 The justification for this is that modal verbs in English act like auxiliaries with respect to questions, negation, as so on. For instance, modals appear to the left of the Neg° head, requiring overt modal-to-T movement in English, whereas most English verbs undergo covert V-to-T.
clauses in German and Dutch, there is a clause-final verb cluster. The rules of ordering the verbs and the morphological form of the verbs within the cluster vary from dialect to dialect. Standard German and standard Dutch are not atypical. Consider the following examples from Kroch and Santorini (1991: ex1&2):

(21) a. daß Hans Peter Marie schwimmen lassen sah  (German)
that Hans Peter Marie to-swim to-let saw
*that Hans saw Peter make Marie swim*

b. daß Hans Peter Marie zu schwimmen zu zwingen verbot
that Hans Peter Marie TO to-swim TO to-force forbade
*that Hans forbade Peter to force Mary to swim.*

(22) a. dat Jan Piet Marie zag laten zwemmen  (Dutch)
that Hans Peter Marie to-swim to-let saw
*that Hans saw Peter make Marie swim*

b. dat Jan Piet Marie verbod te dwingen te zwemmen
that Hans Peter Marie forbade TO to-swim TO to-force
*that Hans forbade Peter to force Mary to swim.*

Also, recall the subordinate clause of (2), repeated here:

(23) Ich glaube nicht, daß er die Lieder wird haben singen können
I believe not that he the songs will to-have to-sing to-be-able
*I do not believe that he will have been able to sing the songs.*

The rules concerning clause-final verb clusters in German or Dutch are too complex to state here easily—let alone the rules which vary from dialect to dialect. However,

These clusters deviate from being head-final or head-initial, which is a motivation for treating them as an instance of overt head-movement. This allows the clusters to be constrained by morphological requirements
reconciling the selectional features of the verbs with their feature-checking requirements, and with their observed theta-role assignment properties, it is clear that the only way to reconcile the data from West Germanic languages with the Minimalist Program is by assuming that modals (and other kinds of verbs) undergo overt head movement—that is, the overt, pre-Spell-Out version of the covert head movement we argued was involved in the derivation of modals in English. This is a direct corroboration of our analysis of modals, as well as the general approach to Amalgamation and Licensing for Visibility.

Moreover, this analysis of modal verbs is precisely analogous to analysis of raising and control structures which were discussed in chapter 2. In the case of modals, our analysis of Case-checking and theta-role assignment allows us to explain the basic properties of modals in a simple way, without the complications of having modal verb complements be either raising or control constructions. Also, this analysis derives the fact that the head of the complement of modal verbs must incorporate into a higher projection with the modal, either overtly or covertly. In fact, this process is observed overtly in many of the West Germanic languages, among others; it is a short step to argue that this process is a linguistic universal, which happens after Spell-Out if it cannot happen before.

3.5 Causatives

In this section, we will look at the structure of causatives in a variety of languages. We will pay particular attention to the analysis of causatives given in Baker (1988) and the criticism of that analysis in Alsina (1992). In light of this work, a theory will be proposed, based on the framework of chapter 2 and this chapter, which explains the properties of causative constructions cross-linguistically.

The first basic fact about causatives is that they take three arguments: a “causer”, a at PF, which seems to be the relevant level for dealing with such string-specific data.

9In fact, the ambiguity between raising and control interpretations has been addressed before. For instance, there has been some discussion of “quasi-arguments” in weather-predicates or in certain raising constructions. (See Chomsky 1981:325 for some discussion.) Also see Pollard & Sag (1994) for discussion of “coercion” of control verb theta assignment.
complement clause, and a “causee” which is also an argument of the complement clause. 
The last argument has often not been recognized as a separate theta role assigned by the 
causative verb, but it should be, as Alsina (1992) argues strongly. For example, consider 
the following sentences:

(24)  a. Andy makes Jamie kiss Alex.
b. Andy makes Alex be kissed by Jamie.

These sentences (24a) and (24b) do not seem to be truth-conditionally synonymous, as 
we saw with the sentences in (1) and (2) in chapter 2, repeated below.

(25)  a. I believe [the doctor \textsubscript{1} to have examined Sam \textsubscript{2}]
b. I believe [Sam \textsubscript{2} to have been examined t\textsubscript{2} by the doctor \textsubscript{1}]

(26)  a. I persuaded the doctor \textsubscript{1} [PRO \textsubscript{1} to examine Sam \textsubscript{2}]
b. I persuaded Sam \textsubscript{2} [PRO \textsubscript{2} to be examined t\textsubscript{2} by the doctor \textsubscript{1}]

We argued that the truth-conditional synonymy of (25a&b) indicated that the matrix verb 
\textit{believe} does not assign a theta role to the subject of the embedded clause. However, the 
difference in the truth-conditional meanings of (26a) and (26b) showed that the matrix 
verb \textit{persuade} in (26) assigns a theta role to the subject of the embedded clause. Since 
there is a lack of synonymy in (24a&b) similar in nature to the lack of synonymy in 
(26a&b), this indicates that the matrix verb in a causative like (24) is a object control 
verb like \textit{persuade}—or in terms of this thesis, it assigns a theta role to the subject of the 
complement clause—and not an ECM verb like \textit{believe}, for which a pair of sentences like 
(21) would be truth-conditionally synonymous. Such a possibility—that causatives are 
object control verbs—is suggested by Alsina (1992) as a possible remedy for the faults in 
the analysis of causatives in Baker (1988), whose analysis assumes that causatives are a
two-place predicate.

Causatives have been an active area of inquiry in linguistics, and there have been a number of different analyses of their structure. Let us assume following Baker (1988) that the causatives in (24), as well as causative constructions universally, take CP complements. Actually it is not critical that this projection is a CP per se. Presumably, it could be any functional projection consistent with the selectional features and the Case features of the causative verb and allowed both the embedded verb and the causative verb to assign theta roles to the subject of the embedded clause. However, we follow Baker (1988) in supposing it a CP.\footnote{We assume this mostly to have a concrete analysis from which to start. In fact, the details of the functional projections associated with the embedded clause in causative constructions is not terribly important to the analysis—unless it makes an analysis impossible, as would happen if there were direct adjunction of the embedded theta-role assigning verb to the causative, which would violate Amalgamation.}

Let us assume following Baker (1988) that the causatives in (24), and in fact causatives in all languages, take CP complements. The $\text{C}^\circ$ head of this CP complement has V-checking features which must be satisfied in one of two ways: either by adjunction of the embedded “caused” verb to $\text{C}^\circ$, or by moving the entire VP to SpecCP—which are the only ways that a verb can move into the checking domain of a $\text{C}^\circ$. Aside from this V-feature checking in the embedded CP, the incorporation of the embedded V into the causative clause is required by the grammar. In (24), this means that the head of the complement (\textit{kiss} or \textit{be kissed}) must move after Spell-Out first to $\text{C}^\circ$ and then to the causative verb in the matrix clause, so that the complement clause may be licensed at LF.

Such movement is overt in many languages. In languages with overt incorporation of the causative, the causative is a morphological affix and therefore the embedded verb must be moved to adjoin to the causative before Spell-Out. Languages like this include Chichewa (discussed later in this section), Turkish (Aissen 1974), Malayalam (Mohanan 1982), Japanese (Baker 1988), and many other languages. In addition, overt movement has been suggested as an analysis for languages without explicit univerbation of the embedded verb and the causative, especially for the causative construction in the Romance
languages (e.g. Burzio 1986, Zubizarreta 1985). Therefore, positing the movement as universal either before or after Spell-Out is natural in the Minimalist Program.

For example, consider in (27) the LF structure for (24a):

(27)

The embedded verb *kiss* undergoes a series of covert (post-Spell-Out) movements: it adjoins to *AgrO* as usual. Then *AgrO* moves to *I*’, allowing as usual the selectional features of all the functional heads to be checked. Then *I*’ moves to *C*’ (also to check selectional features), and then *C*’ moves to the causative verb complex head—again
because of selectional features but also to be Visible at LF for theta-role assignment by the causative verb. At each of these stages the other relevant features—Agreement and Case—are checked. Thus we derive the complex head adjunction structure in (27) at LF.

The Case-chain \{[DP Jamie], t’, t\} can be assigned theta roles by both the embedded verb *kiss* and by the causative verb *made*, and it is Superordinate to the Case-chain \{[DP Alex], t\}, which remains in the embedded clause and hence can only be assigned a theta role by *kiss*. The Case-chain \{[DP Andy], t\} starts in the causative clause, and by ending up in the matrix SpecIP is Superordinate to both. Therefore, the agent theta role of the causative (the “causer”) is assigned to \{[DP Andy], t\}; both the patient theta role of the causative (the “causee”) and the agent theta role of *kiss* is assigned to \{[DP Jamie], t’, t\}; and the patient theta role of *kiss* is assigned to \{[DP Alex], t\}. Finally, the complement CP clause, being Visible, can be assigned the appropriate theta role by the causative verb.

Once again, as was the case with our analysis of modal verbs, we argue that English causative constructions involve a series of unobservable, post-Spell-Out head movements. Of course, this analysis would be more plausible if we could argue that such an analysis is universal, and that there are other languages in which this pattern of head movement is overt. Fortunately, there are such languages, as we see below.

As mentioned above, this analysis is essentially given in Alsina (1992) as a possible control analysis of causatives within GB Theory. However, Alsina argues that it has a number of failings. Based on data from Chichewa, Alsina argues that constraints associated with both theta roles affect the syntactic properties of the “overt” causee object of the causative, which in a “control” analysis of causatives would not explain, because the embedded verb’s theta role is assigned to PRO. However, in the analysis of control constructions given in this work, both theta roles are assigned to the same Case-chain and hence it would be expected that both the theta roles could affect the syntactic properties of the causee object. Also, Alsina argues that given a “control” analysis of causatives, the causee must be identical in all languages with the subject of the embedded clause.
However, this is not true. There are two general types of causatives: in the first, (Alsina’s (1992) variant 1) as we saw above, the causee and the subject of the embedded clause are identical, but in the other type (variant 2), the causee is identical with the object of the embedded clause. The latter variant is impossible to account for in a version of GB Theory or Minimalism which assumes the standard Control Theory and Theta Theory, because there is absolutely no way for the object of the embedded clause either (a) to be assigned a second theta role by the matrix causative verb, or (b) to be a PRO in object position, since it is both governed (thereby violating a Binding-Theory-based Control Theory) and unable to check null Case (therefore violating a Case-theoretic approach to Control Theory). However, by adapting a suggestion of Baker (1988) we can show that both variants can be accounted for. Variant 1 is shown above, of course, but what about variant 2?

It is known that there are two ways for a head to check a feature of another head: the head can adjoin to the checking head, or the maximal projection of the head can move to the Specifier of the checking head. In variant 1, as we saw above, the verb adjoined to C˚ to check selectional features there. Let us suppose that in some languages this checking can be achieved by the VP moving to SpecCP.

Consider the following sentences in Chichewa, taken from Alsina (1992: ex. 12):

(28) a. Nũngu i-na-phík-íts-a maũngu kwá kádzĩdzi
   9.porcupine 9s-PS-cook-CST-FV 6.pumpkins to 1a.owl
   ‘The porcupine had the pumpkins cooked by the owl.’

b. Nũngu i-na-phík-íts-a kádzĩdzi maũngu
   9.porcupine 9s-PS-cook-CST-FV 1a.owl 6.pumpkins
   ‘The porcupine made the owl cook the pumpkins.’

Although both sentences use the causative morpheme -íts-, there is a difference in meaning
between these sentences. In (28b) we have a sentence which is analogous to (24a), and we assume that its LF is very similar to (27), save for the fact that in Chichewa there is overt incorporation of the verb into the causative head. However, (28a) differs from (28b) in several ways. The first is plainly syntactic: the logical object of the embedded verb is unmarked and adjacent to the verb, while the logical subject of the embedded clause is in a prepositional phrase. There are also semantic differences: in (28a) the causee theta role of the causative morpheme is assigned to the object *maũngu* ‘pumpkins’ of the embedded verb, rather than *kádzĩdzĩ* ‘owl’, the subject of the embedded verb, so (28a) and (28b) are not truth-conditionally synonymous.

Let us suppose that Chichewa allows its causative morpheme to select a CP complement whose head may check verbal features either by head-adjunction or by substitution to SpecCP. If the former choice is made, then sentences analogous to (24a) with LFs like (27) are generated. However, if the VP adjoins to SpecCP instead, there are a number of implication for the rest of the derivation. The first and most important is that the object “leapfrogs” the subject, and this allows the object to move into the higher clause to check its Case features—and in addition it is assigned the causee theta role by the causative morpheme. This effectively blocks the embedded subject from moving to the higher clause to check its Case, and therefore must get its Case within the lower clause. For this, we stipulate that the combination of AgrO° and I′ checks case on the embedded subject, which is expressed by the presence of a preposition—in Chichewa, *kwá*. So for (28a) we would expect a LF like the following:
Once the embedded VP moves to SpecCP, the shortest move to check Case is from within that VP. Therefore, the DP *maûngu* needing Case is extracted from the VP in SpecCP, leaving the DP *kádzćdzi* stranded in SpecIP where it cannot be assigned Case in the usual fashion. However, we hypothesize that *AgrO* when it is licensed by the verb’s Case features but otherwise unused can adjoin to *I* and assign dative Case (which in Chichewa is shown by the presence of the preposition *kwá*). So in (24) there are three
Case-chains: \{[[DP Nüngu], t}, \{[[DP maûngu], t}, \{[[DP kwá kádzdzi], t}. The causative morpheme assigns two theta roles; \{[[DP Nüngu], t} and \{[[DP maûngu], t} are available, and since \{[[DP Nüngu], t} is Superordinate to \{[[DP maûngu], t}, \{[[DP Nüngu], t} is assigned the agent “causer” theta role and \{[[DP maûngu], t} is assigned the patient “causee” theta role.

The embedded verb -phik- “cook” also assigns two theta roles, and to it the Case-chains \{[[DP maûngu], t} and \{[[DP kwá kádzdzi], t} are available. Therefore \{[[DP maûngu], t} is assigned the agent theta role of phik and \{[[DP kwá kádzdzi], t} the patient theta role.11

From these examples, it is evident that the theory presented here correctly captures the correct generalizations about causatives in Baker (1988) and Alsina (1992) in a Minimalist framework with very little further work. The Case features checked by the various arguments of the embedded clause, whether transitive or intransitive, are derived in a way which is a rough extension of Baker (1988) to Minimalism; the binding properties of causatives would also accounted for in the fashion of Baker (1988), given that our theory proposes the same kinds of structures proposed there. Finally, the different types of causatives are related directly to the two different kinds of features checking: head adjunction and movement to a specifier. The advantage of the analysis presented here is that we can also account for the properties of causatives discussed in Alsina (1992), which require that more than one theta role be assigned to the same Case-chain.

11 An attentive reader will notice that, given the definition of Superordinate in §2.2, the theta roles of -phik- “cook” should be assigned the other way around, because the Case-chain \{[[DP maûngu], t} is Superordinate to the Case-chain \{[[DP kwá kádzdzi], t}, yielding the wrong result. Fortunately, there is a simple and reasonable way out of this problem. As defined in §2.2, the definition of Superordinate looks over arbitrarily large amounts of syntactic structure comparing Case-chains to find the maximally c-commanding position within those Case-chains. However, such a “global” capacity is undesirable in the Minimalist Program, and we can change this so that the definition of Superordinate only takes into account the Complete Functional Complex (CFC) of the theta-assigner in question. So the revised definition is as follows:

(i) \textbf{Superordinate}

Given two chains \(\alpha\) and \(\beta\) and a theta-assigning head \(\theta\), \(\alpha\) is Superordinate to \(\beta\) with respect to \(\theta\) iff some position in \(\alpha\) which is in the CFC of \(\theta\) c-commands every position in \(\beta\) which is in the CFC of \(\theta\).

This slight redefinition of Superordinate captures the correct generalization here without losing any of its previous empirical successes.
3.6 Conclusions

In this chapter, we have considered three constructions: small clauses, modal verbs and causatives, each of which posed a potential problem for the theory presented in this thesis. In each of these cases, we saw that the theory developed in chapter 2, combined with a simple theory specifying the relation between theta role assignment and head movement could explain the properties of these constructions in a simple and clear manner.

In this analysis, small clause constructions, modal verb constructions and causative constructions all share the requirement that the head of their embedded clause must adjoin to the head of the clause which selects the embedded clause. This satisfies the selections features of the heads, and also allows the embedded clause to be Visible at LF for theta-role assignment. On the other hand, Amalgamation restricts the ability of two heads to both assign a theta role to the same Case-chain. The interaction of these requirements entails that control is impossible with small clauses, and entails the ambiguity between the root and epistemic interpretations of modal verbs. Finally, it also offered an analysis of causatives which is arguably better than the standard approach.

However, both this chapter and chapter 2 were largely concerned with properties of English constructions. In the next chapter, we will look at control in a variety of Romance and Germanic languages; this will allow us to consider control constructions, and the theory presented in this thesis, from a comparative point of view.