1 *Introduction*

In the following pages we develop an extended argument for a proposal whose conceptual simplicity and empirical success will, we trust, be evident to all readers. The proposal says that (obligatory) control is movement, more specifically, A-movement. We propose that the phenomena that have been used to motivate a special and separate control construction are best explained if control is treated as an A-movement dependency, on a par with other phenomena that have been traditionally treated in terms of A-movement such as passive, raising, and (local) scrambling. Put another way, we claim that maintaining the constructional specificity of control (in whatever form, be it in terms of the PRO theorem [e.g., Chomsky 1981], null case [e.g., Chomsky and Lasnik 1993; Martin 1996; and Bošković 1997], or *ad hoc* "anaphoric" tense-agreement dependencies [e.g., Landau 1999, 2000, 2004]) significantly hampers our understanding of the phenomenon as it leads to explanations that are roughly as complex as the phenomenon itself.

Despite virtues that we believe are transparent (see e.g., Hornstein 1999, 2001), the movement theory of control (hereafter, MTC) has proven to be quite controversial.[1] We believe that there are several reasons for this. The first one is historical. Differentiating raising from control in terms of movement has been a fixed point within generative grammar from the earliest accounts within the standard theory to current versions of minimalism (see Davies and Dubinsky 2004). Under this long-held view, which became crystallized in GB with the formulation of the (construction-specific) control module (Chomsky 1981), if raising involves movement, control cannot. It is thus not surprising that the MTC has been welcomed with considerable skepticism, as its basic proposal is exactly to analyze control in terms of (A-)movement. However, such historical bias should not deter us from a fair evaluation of the conceptual properties and empirical coverage of the MTC.

1 See e.g., Landau (2000, 2003); Culicover and Jackendoff (2001, 2005); Kiss (2005); and van Craenenbroeck, Rooryck, and van den Wyngaerd (2005) for a useful sample.

The second reason behind the controversy is also related to the long interest control has enjoyed within the generative tradition. Over the years, control phenomena have been richly described. Consequently, any new approach will likely fail, at least initially, to adequately handle some of the relevant data. Moreover, if the novel approach is conceptually tighter than the more descriptive accounts that it aims to replace (as we believe to be the case with the MTC), some features of the phenomenon heretofore assumed to be central may not be accommodated at all. This should occasion no surprise, as it reflects the well-known tension between description and explanation. Odd as it may seem, failure to cover a data point may be a mark of progress if those that are covered follow in a more principled fashion. The virtues of a proposal can be seriously misevaluated unless one keeps score of both what facts are covered and *how* facts are explained. A weak theory can often be "easily" extended to accommodate yet another data point, and this is *not* a virtue. Correspondingly, a tight theory may miss some "facts" and this is not necessarily a vice, particularly if the account is comparatively recent and the full implications of its resources have not yet been fully developed. We believe that many have been too impressed by these apparent problems without considering how the MTC might be developed to handle them. In fact, we believe that the MTC actually faces few empirical difficulties (and none of principle), whereas the current alternatives both face very serious empirical hurdles (e.g., backward control) and often empirically succeed by stipulating what should be explained (e.g., the distribution of PRO through null case). One aim of what follows is to make this case in detail.

Finally, it is fair to say that the resistance to MTC is in part due to the inadequacies and limitations of previous versions of the MTC (including our own work), which we have tried to overcome here. Addressing the vigorous critiques of MTC here and in previous work (Hornstein 2003; Boeckx and Hornstein 2003, 2004, 2006a; Nunes 2007; Boeckx, Hornstein, and Nunes in press) has allowed us to rectify some errors, clarify the proposal, and sharpen the arguments. This stimulating intellectual exercise has led us to better appreciate the consequences of the MTC and has in fact convinced us that it covers even more empirical ground than we at first thought, as we will argue in the following chapters.

For all these reasons, we thought that a detailed defense of MTC required a monograph. But before we launch our defense of MTC, a few notes are in order.

First, we cannot emphasize enough that MTC does *not* equate "control" with "raising." Since the MTC was first proposed, it has been regularly objected

that the MTC cannot be right because of features that control has, but raising does not, and vice versa. However, control is raising only in the descriptive sense that control is an instance of A-movement, but it is not raising *qua* construction. In other words, all the MTC is saying is that, like the derivation of raising, passive, or local scrambling constructions, the derivation of obligatory-control constructions also involves A-movement. The different properties of constructions involving *wh*-movement and topicalization, for instance, do not argue against analyzing them in terms of A'-movement. Similarly, we urge the reader not to dismiss our proposal simply because (unanalyzed) control–raising asymmetries exist. Although raising often proves useful in illustrating properties of A-movement that carry over to control, it is a ladder that ought to be kicked away as theory advances. In the chapters that follow, we in fact argue that control–raising asymmetries generally reduce to independent factors – something we take to be an indication that the MTC is on the right track.

Second, the MTC is actually not a radically new idea. It goes back as far as Bowers (1973), who already proposed that raising and control should be basically generated in the same way. However, as the proposal conflicted with core principles of almost every model of UG from *Aspects* to GB, it did not find fertile soil to blossom for a long time. This scenario drastically changed when the minimalist program came into the picture. Chomsky's (1993) proposal that D-structure should be eliminated provided a very natural conceptual niche for the MTC within the generative enterprise as it removed the major theoretical obstacle that prevented movement to θ-positions. In a system with D-structure, movement to θ-positions is a non-issue, for movement can only take place once θ-assignment is taken care of. By contrast, in a system without D-structure, where movement and θ-assignment intersperse, movement to θ-positions arises at least as a logical possibility. Thus, whether or not it is a sound option has to be determined on the basis of the other architectural features of the system, as well as its empirical coverage. We hope to show that the MTC fits snugly with some leading minimalist conceptions and thus constitutes an interesting argument in its favor.

Third, as minimalism aspires to *explain* why UG properties are the way they are, we are interested in developing a theory of control that *deduces* the properties of control configurations from more basic postulates, rather than merely listing the possible controllers, controllees, control predicates, and control complements coded as features of individual lexical items.

Finally, although our specific implementation of MTC is the one that has been extended to the broadest range of data thus far, it is certainly not the only one possible. O'Neil (1995), Manzini and Roussou (2000), Kayne (2002), and

Bowers (2006) share the spirit but not the details of our analysis. For reasons of space, we will not be able to do proper justice to these works and the reader is invited to evaluate each different implementation in its own right.

Let us close this introductory chapter by providing an overview of the subsequent chapters. Chapter 2 offers a brief overview of how control is handled in the standard-theory framework, in GB, and in non-movement approaches within minimalism. Chapter 3 lays out the broad features of our version of the MTC. Chapter 4 discusses some of the empirical advantages that the MTC has. Chapter 5 addresses many of the empirical challenges that have been considered to be fatal to the MTC and proposes solutions compatible with the MTC. Chapter 6 presents our take on how non-obligatory control is to be analyzed. Chapter 7 discusses the extent to which the MTC is based on more solid conceptual and empirical grounds than semantic/selectional approaches to obligatory control. Finally, Chapter 8 concludes the monograph.

2 *Some historical background*

2.1 Introduction

Up to very recently, there had been a more or less uncontroversial view that control phenomena should be analyzed in terms of special grammatical primitives (e.g., PRO) and construction-specific interpretive systems (e.g., the control module). In this chapter, we examine how this conception of control was instantiated in the standard-theory framework (section 2.3), in GB (section 2.4), and in non-movement analyses within the minimalist program (section 2.5), briefly outlining what we take to be the virtues and problems of each approach.[1] This discussion will provide the general background for us to discuss the core properties of (our version of) the MTC in Chapter 3 and evaluate its adequacy in the face of the general desiderata for grammatical downsizing explored in the minimalist program.

2.2 What any theory of control should account for

A theoretically sound approach to control – one that goes beyond the mere listing of the properties involved in control – must meet (at least) the following four requirements.

First, it must specify the kinds of control structures that are made available by UG and explain how and why they differ. Assuming, for instance, that obligatory control (OC) and non-obligatory control (NOC) are different, their differences should be reduced to more basic properties of the system.

Second, it must correctly describe the configurational properties of control, accounting for the positions that the controller and the controllee can occupy. In addition, it should provide an account as to why the controller and the controllee are so configured. Assuming, for instance, that the controllee can

1 For much more detailed discussion, we urge the reader to consult Davies and Dubinsky's (2004) excellent history of generative treatments of raising and control.

only appear in a subset of possible positions (e.g., ungoverned subjects), why are controllees so restricted?

Third, it must account for the interpretation of the controllee, explaining how the antecedent of the controllee is determined and specifying what kind of anaphoric relation obtains between the controllee and its antecedent (in both OC and NOC constructions) and why these relations obtain and not others. For instance, assuming that controllers must locally bind controllees in OC constructions, why is the control relation so restricted in these cases?

Fourth, it must specify the nature of the controllee: what is its place among the inventory of null expressions provided by UG? Is it a formative special to control constructions or is it something that is independently attested?

In the next sections, we briefly review how these concerns have been addressed from the standard-theory model to the minimalist program.

2.3 Control in the standard-theory framework

Within the framework of the standard theory, control phenomena were coded in the obligatory transformation referred to as *equi(valent) NP deletion* (*END*), which for our current purposes can be described as follows:[2]

(1)

	X-NP-Y-[$_S$	{for/poss}-	NP-	Z]-	W			
Structural description:	1	2	3	4	5	6	7	→
Structural change:	1	2	3	4	Ø	6	7	

Conditions: i. 2 = 5
ii. the minimal-distance principle is satisfied

Irrelevant details aside, END applies to the (a)-structures in (2)–(5), for instance, and converts them in the corresponding (b)-sentences.

(2) a. **John** tried/wanted/hoped [for **John** to leave early] →
b. John tried/wanted/hoped to leave early

(3) a. **John** regrets/insisted on/prefers [poss **John** leaving early] →
b. John regrets/insisted on/prefers leaving early

(4) a. John persuaded/ordered/forced/asked/told **Mary** [for **Mary** to leave early] →
b. John persuaded/ordered/forced/asked/told Mary to leave early

(5) a. **John** kissed Mary before/after/without [poss **John** asking if he could] →
b. John kissed Mary before/after/without asking if he could

2 Here we abstract away from issues that are orthogonal to our discussion such as the interaction between END and the rule of complementizer deletion, which has the effect of deleting the term numbered 4 in (1). See Rosenbaum (1967, 1970) for discussion.

According to this approach, there is nothing of special interest in the nature of the controllee. It is a regular NP in the underlying structure and the fact that the corresponding surface position is phonetically null follows from the kind of transformation END is. It is a deletion transformation that removes the targeted NP, leaving nothing at surface structure. To put it differently, the superficial phonetic difference between controller and controllee results not from intrinsic lexical properties of the controllee, but from properties of the computation itself, i.e., that END is a deletion operation.

As far as the configurational properties of control are concerned, END explicitly specifies that the controllee (the target of deletion) must occur in the subject position of infinitival clauses (*for*-clauses) and gerunds (*poss*-clauses), and that the controller must be the closest NP (in compliance with the minimal-distance principle). Thus, according to the minimal-distance principle, sentences such as (4b) must be derived from the structures in (4a) and not from the one in (6) below, which would incorrectly allow the understood subject of the embedded clause to be interpreted as being coreferential with the matrix subject. As opposed to what we find in (4a), the antecedent of the controllee in (6) is not the closest NP around. As for adjunct control in sentences such as (5), the minimal-distance principle is satisfied under the assumption that the embedded clause is adjoined to the matrix clause and, as such, it is structurally closer to the subject than it is to the object.[3]

(6) **John** persuaded/ordered/forced/asked/told Mary [for **John** to leave early]

Finally, the interpretation properties of control are enforced by condition (i), which requires that controller and controllee be "identical," which was understood in terms of coreference.

This general approach was refined within the standard theory as more complex control structures were considered, but its axiomatic (i.e., stipulative) nature remained. The configurational and interpretive properties of control were analyzed as irreducible features of the END transformation itself. This by no means diminishes the value of these earlier approaches to control. Identifying the different properties of control phenomena with such formal rigor

3 END as stated is not entirely adequate empirically. Given (1) above, the structure in (ia), for example, should allow for control by 'Mary' in (ib):

(i) a. John persuaded a friend of **Mary** [for **Mary** to leave]
b. John persuaded a friend of Mary to leave

It should be clear how requiring that some sort of command relation hold between the antecedent NP and the deleted one will help screen out cases like (i), where the "wrong" NP is chosen.

was unquestionably an achievement, with large consequences for theorizing beyond control structures, and it paved the way for subsequent reanalyses in GB and in the minimalist program.

Before we leave this brief review, two points are worth mentioning which will be relevant to the discussion of these later reanalyses, including the MTC. The first one regards an empirical problem that the standard-theory approach faced in relation to the way it handled the interpretive properties of control. As we saw above, the controller and the controllee were taken to be lexically identical and the semantic relation between them was understood as coreference. Problems arise when the controller is not a referential NP, as exemplified by the contrast between (7) and (8).

(7) a. [**John** wants [**John** to win]] →
b. John wants to win

(8) a. [**Everyone** wants [**everyone** to win]] →
b. Everyone wants to win

Whereas (7a) might be taken to roughly represent the meaning of (7b), (8a) in no way represents the interpretation of (8b), which should rather be paraphrased as 'Everyone wants himself to win.' This suggests that, instead of an NP identical to (i.e., coreferential with) its controller in underlying structure, what we actually need is a kind of bound anaphor or an expression that can be so interpreted.[4] The obvious question then is how to obtain this bound interpretation.

The second point worth mentioning concerns the identification of another type of control. Relatively early on, END was distinguished from a related operation dubbed super-equi (SEND). This operation also deletes a subject of a non-finite clause but, in contrast to END, it operates across unbounded stretches of sentential material, as illustrated in (9).[5]

4 If there is an anaphoric relation in control structures, then END is unlikely to be a chopping ("gap"-leaving) rule. Rather, it is more like the rules of reflexivization or pronominalization, which were operations governed by command relations. The problem is that control structures do not *appear* to leave lexical residues like the other construal operations. They appear to require a phonetic gap. Seen from a contemporary perspective, the problem of how to characterize the rules that lead to control structures (are they chopping rules or construal rules?) highlights the tension that we will see constantly recurring: how best to account for both the distribution of the controllee and its interpretation.

5 See Grinder (1970). Data such as (9) are not the sorts of cases Grinder discussed, but they fall under the SEND rubric.

(9) a. [$_{S1}$ **John** said [$_{S2}$ that Mary believes [$_{S3}$ that [$_{S4}$ **John** washing himself] would make a good impression on possible employers]]] →
b. John said that Mary believes that washing himself would make a good impression on possible employers

Note that (9) violates the minimal-distance principle, as 'Mary' intervenes between the target of deletion ('John' in S_4) and its antecedent ('John' in S_1). Moreover, in contrast to standard END configurations, the controllee is not within a clausal complement (or adjunct) of a higher predicate. In (9), for instance, the controllee is within the sentential subject of S_3. The following question then arises: what is the relation between END and SEND? Or to put the question somewhat differently: why should UG have two rules that have the same effect (deletion of an identical NP), but apply to different configurations?[6]

In the next sections we examine some answers to these two issues that were offered within GB and the minimalist program.

2.4 Control in GB

Building on earlier work in the extended standard theory (EST), the GB approach to control is considerably more ambitious and empirically more successful than the standard-theory model.

Within GB, the controllee is a PRO, a base-generated NP containing no lexical material ($[_{NP}\ \emptyset]$). This conception of the controllee as a base-generated non-lexical formative arises as a natural consequence of the GB assumptions regarding the base component. The GB theory of the base includes both phrase-structure rules, like the ones in (10), and lexical-insertion operations, like the ones in (11).

(10) a. S → NP INFL VP
b. VP → V NP
c. NP → N

(11) a. N → John/he/it/Bill
b. V → kiss/see/admire
c. INFL → past/to

These two types of rules operate in tandem to generate structures such as (12) below. However, they can also be used to generate structures like (13), where the subject of the clause has been generated by the phrase-structure component

6 Grinder (1970) actually collapsed END and SEND. However, later approaches identified many substantial differences between the constructions underlying END and SEND that are better captured if two kinds of control are recognized, as we shall see below.

but has not been filled by lexical insertion. In short, a theory of the base factored into a set of phrase-structure rules and lexical-insertion operations has room for an element like PRO: it is what one gets when one generates an NP structure but does not subject it to lexical insertion.

(12) $[_S\ [_{NP}$ John] past $[_{VP}$ see $[_{NP}$ Bill]]]

(13) $[_S\ [_{NP}\ \emptyset]$ to $[_{VP}$ see $[_{NP}$ Bill]]]

This way of understanding PRO has an interesting consequence for the constructions that were captured by END in the standard theory. If one assumes that categories without lexical content are uninterpretable unless provided with "content" (by being linked with an antecedent, for example) and, furthermore, that the principle of full interpretation does not tolerate contentless structures, then the requirement that PRO must have an antecedent follows naturally.[7]

We wish to stress this point as it is important for some of the discussion that follows. If one treats PRO as a lexical element, it is hard to explain *why* PRO must be phonetically null and *why* it requires an antecedent. Of course, it is possible to stipulate that these two features are inherent properties of a specific lexical item (PRO), but this cannot explain *why* PRO is necessarily anaphoric and null. Moreover, so conceived, PRO is a rather unusual lexical element as it has no positive properties. It has no phonetic matrix and its only semantic feature is the requirement that it must be coindexed with a *grammatical* antecedent.[8] This point is worth emphasizing. PRO, on this view, is not simply a semantically dependent expression that needs to be interpreted with respect to some salient element in the discourse (e.g., like 'the other' in 'John ate one of the bagels. Harry ate the other.'). Rather, PRO is specified as needing an antecedent in a particular *structural* configuration. However, this is a very odd *lexical* feature as it is only definable in configurational (i.e., grammatical) terms. In other words, invoking such features in the construction of lexical items (be it PRO or any other item) is just a way of simulating a grammatical requirement via lexical stipulation.[9]

The GB approach offers a sounder alternative as it treats PRO's properties as the result of interacting grammatical principles. This feature of the GB analysis

7 See Chomsky (1980: 8): "If Coindex does not apply and the embedded clause contains PRO, then we end up with a 'free variable' in LF; an improper representation, not a sentence but an open sentence."

8 This point is similar to Chomsky's (1995) argument against considering Agr as a lexical category. Given that its only features are uninterpretable, a preferable approach, all things being equal, is to take these features as belonging to related true lexical categories.

9 For a discussion of reflexives and bound pronouns in light of this discussion, see Hornstein (2001, 2007).