



# Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

## Introduction

Broekhuis, H.; Vogel, R.

### **published in**

Linguistic Derivations and Filtering. Minimalism and Optimality Theory  
2013

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in KNAW Research Portal](#)

### **citation for published version (APA)**

Broekhuis, H., & Vogel, R. (2013). Introduction. In H. Broekhuis, & R. Vogel (Eds.), *Linguistic Derivations and Filtering. Minimalism and Optimality Theory* (pp. 1-28). (Advances in Optimality Theory). Equinox Publishing Ltd.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[pure@knaw.nl](mailto:pure@knaw.nl)

# 1 Introduction

Hans Broekhuis\* and Ralf Vogel\*\*

**ABSTRACT** This chapter will motivate why it is useful to consider the topic of derivations and filtering in more detail. We will argue against the popular belief that the minimalist program and optimality theory are incompatible theories in that the former places the explanatory burden on the generative device (the computational system  $C_{HL}$ ) whereas the latter places it on the filtering device (the OT evaluator). Although this belief may be correct in as far as it describes existing *tendencies*, we will argue that minimalist and optimality theoretic approaches normally adopt more or less the same global architecture of grammar: both assume that a generator defines a set  $S$  of potentially well-formed expressions that can be generated on the basis of a given input and that there is an evaluator that selects the expressions from  $S$  that are actually grammatical in a given language  $L$ . For this reason, we believe that it has a high priority to investigate the role of the two components in more detail in the hope that this will provide a better understanding of the differences and similarities between the two approaches. We will conclude this introduction with a brief review of the studies collected in this book.

## 1. The architecture of grammar

The studies collected in this book all discuss the relation between the generative and the filter component of the grammar. The focus will be on syntax although the collection also contains a contribution by John J. McCarthy and Kathryn Pruitt, which discusses the issue for phonology. The starting point of this book is the popular view that current generative theories differ considerably in where they place the burden of explanation: whereas minimalist approaches generally assume that this is the generative component (the computational system  $C_{HL}$ ), optimality-theoretic approaches generally focus on the filter component (the OT-evaluator). This difference between the minimalist program (MP) and optimality theory (OT) is also reflected in the claims that are normally made about the output of the generator;

---

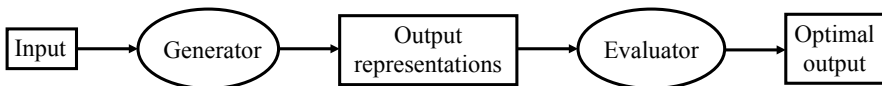
\* Hans Broekhuis, Meertens Institute, P.O.-box 94264, 1090 GG Amsterdam, The Netherlands. E-mail: hans.broekhuis@meertens.knaw.nl

\*\* Ralf Vogel, Universität Bielefeld, Fakultät für Linguistik und Literaturwissenschaft, P.O.-Box 10 01 31, 33501 Bielefeld, Germany. E-mail: rvogel@uni-bielefeld.de)

minimalist approaches normally presuppose that the output of  $C_{HL}$  is small and may in fact be restricted to a single representation in many cases; optimality-theoretic approaches, on the other hand, normally maintain that the generator creates a candidate set that is very large or even infinite. It is important to note, however, that proponents of MP normally accept the idea that the generator may overgenerate and that we must therefore assume additional means to filter out the unwanted structures from the reference set. This means that many proponents of MP and OT do agree that the global architecture of grammar has the form in Figure 1.1, where the Generator and the Evaluator can be held responsible for respectively the universal and language-specific properties of languages. The essential property of this model is that the generator defines a set  $S$  of potentially well-formed expressions that can be generated on the basis of a given input, and that the evaluator selects those expressions from  $S$  that are actually grammatical in a given language  $L$ .

This general idea is, of course, not new and has already been formulated by Chomsky and Lasnik in ‘Filters and control’ (1977), where it is argued that ‘to attain explanatory adequacy it is in general necessary to restrict the class of possible grammars, whereas the pursuit of descriptive adequacy often seems to require elaborating the mechanisms available and thus extending the class of possible grammars’. In order to solve this tension they propose that ‘there is a theory of core grammar with highly restricted options, limited expressive power, and a few parameters’ next to a more peripheral system of ‘added properties of grammar’, which ‘we may think of as the syntactic analogue of irregular verbs’. Chomsky and Lasnik assume that core grammar consists of the phrase structure and transformational rules (the generator in Figure 1.1), whereas the more peripheral system consists of language-specific surface filters (the evaluator), and claim that the introduction of these filters contributes to the simplification of the transformational rules by bearing ‘the burden of accounting for constraints which, in the earlier and far richer theory, were expressed in statements of ordering and obligatoriness, as well as all contextual dependencies that cannot be formulated in the narrower framework of core grammar’.

The ideas about which aspects of grammar should be considered part of core grammar or part of the periphery have, of course, considerably changed over the years; the *that*-trace filter, for example, was originally proposed as a language-specific filter for English, but the Empty Category Principle, which ultimately grew out of it, was assumed to be part of core grammar. Nevertheless, the gist of the proposal has survived in the more recent minimalist incarnations of the theory, where core syntax can be more or less equated with  $C_{HL}$ , and the periphery with the interface conditions. The task of reducing core grammar as much as possible has been very successful: the reduction of  $C_{HL}$  to its absolute minimum (internal and external merge) much contributes to the explanatory adequacy of the theory in the technical



**Figure 1.1** The architecture of grammar.

sense that it provides a principled basis, independent of any particular language, for the selection of the descriptively adequate grammar of each language (although it still remains to be shown that the minimalist conception of UG is more successful in core explanatory tasks like modeling language acquisition, explaining the origins of language and so on, than its predecessors). But, as expected, the contribution of core grammar to the descriptive adequacy of the theory has diminished accordingly, so that in this respect we have to rely more and more on the interface conditions.

That the global architecture of grammar has been like that indicated in Figure 1.1 for over three decades now may have been obscured by the fact that earlier phases of the theory assumed a so-called T- or inverse Y-model, according to which the derivation of LF- and PF-representations diverge after a certain point (s-structure or Spell-Out) in order to account for certain mismatches between linear order and semantic interpretation by deriving different input structures for the LF- and PF-component of the grammar. We will see in Section 2.1 that this property of the early *principles-and-parameters* (P&P) models has disappeared in the later versions of MP and that, as a result, these later versions fully accord with the linear model in Figure 1.1. Now we have established this, we will briefly discuss the form and the role of the filter component in various stages of the P&P framework, as well as in OT.

## 2. The filter component

Although Chomky and Lasnik originally assumed that the filter component was not part of core grammar but of a language-specific periphery, it soon became clear that the filter component also had certain universal properties. For this reason, we will replace the notion of core grammar for the derivational component by the notions *generator*, *computational system* and *narrow syntax*. In order to avoid unwanted connotation we will likewise avoid the notion of periphery and use instead the notion of *evaluator* or *filter component*. The generator and the evaluator should both be considered part of core grammar.

### 2.1 The *principles-and-parameters* approach

The introduction of a filter component in ‘Filters and control’ was motivated by the fact that this made a more restrictive formulation of narrow syntax possible by eliminating ordering statements and language-specific properties from the transformational component of the grammar. By way of demonstration we will consider the derivation of the relative clauses in (1).

- (1) (a) the man who I know  
 (b) the man that I know  
 (c) the man I know  
 (d) \*the man who that I know

The relative pronoun *who* is base-generated in the regular object position, so that the d-structure of the examples in (1) is as given in (2a). Chomsky and Lasnik further proposed that universal grammar (UG) contains a universal principle ‘Move *wh*-phrase’ that requires that relative pronouns (and other *wh*-phrases) be placed to the left of the complementizer, as in the s-structure representation in (2b).

- (2) (a) the man [that I know who] (d-structure)  
 (b) the man [[<sub>COMP</sub> who that] I know  $t_{who}$ ] (s-structure)

The examples in (1) can now be derived by assuming that UG contains a PF-rule Deletion that precedes the filters and freely deletes the relative pronoun *who* or the complementizer *that*; cf. Chomsky and Lasnik (1977: ex. (6)). The resulting PF-representations are given in (3). The desired grammaticality pattern is derived by postulation of the language-specific Doubly Filled COMP Filter, which prohibits the simultaneous realization of the relative pronoun and the complementizer in English. This excludes representation (3d).

- (3) (a) the man [[<sub>COMP</sub> ~~who that~~] I know  $t_{who}$ ]  
 (b) the man [[<sub>COMP</sub> ~~who~~ that] I know  $t_{who}$ ]  
 (c) the man [[<sub>COMP</sub> ~~who that~~] I know  $t_{who}$ ]  
 (d) \*the man [[<sub>COMP</sub> who that] I know  $t_{who}$ ]

Although the deletion rule is freely applicable, the resulting representation is subject to a recoverability principle, which requires that deleted elements be locally recoverable. This is needed to block deletion of the *wh*-phrase in representations like (4): the recoverability principle in tandem with the Doubly Filled COMP Filter ensures that the representations in (4b–d) are excluded.

- (4) (a) I wonder [who ~~that~~ you met  $t_{who}$ ]  
 (b) \*I wonder [~~who~~ that you met  $t_{who}$ ]  
 (c) \*I wonder [~~who that~~ you met  $t_{who}$ ]  
 (d) \*I wonder [who that you met  $t_{who}$ ]

By the same means, deletion of a preposed PP in relative clauses like (5) is blocked. Deletion of *about which* would violate the recoverability principle because the preposition *about* cannot be recovered locally.

- (5) (a) the book [about which ~~that~~ he spoke  $t_{about\ which}$ ]  
 (b) \*the book [~~about which~~ that he spoke  $t_{about\ which}$ ]  
 (c) \*the book [~~about which that~~ he spoke  $t_{about\ which}$ ]  
 (d) \*the book [about which that he spoke  $t_{about\ which}$ ]

The virtue of Chomsky and Lasnik’s proposal is that by accounting for the language-particular properties of English relative constructions by means of the Doubly Filled COMP Filter, we can keep the transformational rule that derives s-structure (2b) maximally simple (Move *wh*-phrase), which, in turn, makes it possible to attribute this rule to UG.

In the *Government and Binding* (Chomsky 1981) and *Barriers* (Chomsky 1986) period, the attempts to further reduce the transformational component of narrow syntax led to the formulation of the general rule Move  $\alpha$ . As far as the filter component was concerned, it turned out that some of the language-specific filters proposed in Chomsky and Lasnik (1977) had a wider application and could be reformulated as more general principles. For example, the so-called *that*-trace filter, which prohibits a trace immediately to the right of the complementizer *that*, was reformulated as/reduced to the Empty Category Principle (ECP), which requires that a trace be properly governed. Although the ECP was claimed to be universal, that is, to be part of UG, its function is more or less the same as that of the *that*-trace filter: it excludes structures that have been created by narrow syntax. Therefore the formulation of the ECP is not a reason to frown with a skeptical eye on the notion of filter; it rather opened the prospect of obtaining a certain degree of explanatory adequacy in the domain of filters, so that the filter component could also enter the domain of core grammar.

In the *Minimalist Program*, as developed by Chomsky since the mid-1980s, the generator seems to have been reduced to its absolute minimum. The computational system of human language  $C_{HL}$ , as it is now called, consists essentially of one merge operation in two guises. External merge combines two independent syntactic objects into a larger syntactic unit, whereas internal merge takes some element from an existing syntactic object, and merges it to the root of this object, thus deriving the effect of movement. Merge is subject to a number of general conditions. For example, it never involves more than two objects at the same time, which results in binary branching phrase structures. Internal Merge obeys certain locality restrictions and is further subject to the Last Resort Condition, which requires that movement be triggered by some unvalued formal feature. As in Chomsky and Lasnik (1977), descriptive adequacy lies mainly outside the computational system: Chomsky (1995: §4.7.3), for example, suggests (rightly or wrongly) that 'rearrangement' phenomena like extraposition, right-node raising, VP-adjunction and scrambling are essentially the result of stylistic rules of the phonological component.

Many of the filters as discussed in Chomsky and Lasnik (1977) have not found an alternative account in MP, but the fact that they are not *discussed* is, of course, no guarantee that they are not *needed*. In this connection it is important to note that Chomsky (1995) explicitly claims that  $C_{HL}$  generates a set of converging (= potentially well-formed) derivations satisfying Full Interpretation, the so-called *reference set*, from which the admissible structures are selected by a number of global economy conditions: derivations with a smaller number of derivational steps are preferred (fewest steps), as are derivations with shorter movement chains (shortest steps).

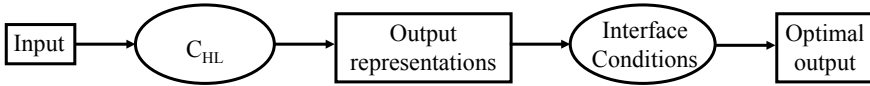
The language  $L$  thus generates three relevant sets of derivations: the set  $D$  of derivations, a subset  $D_C$  of convergent derivations of  $D$ , and a subset  $D_A$  of admissible derivations of  $D$ . FI determines  $D_C$ , and the economy conditions select  $D_A$ . ...  $D_A$  is a subset of  $D_C$ .

(Chomsky 1995: 220)

It is not so clear whether global economy conditions still play a role in the current versions of MP. It seems that very soon they lost independent status by being incorporated into the definition of the movement operation: fewest steps was replaced by Last Resort (Chomsky 1995: 280) and shortest steps by the Phase Impenetrability Condition proposed in Chomsky (2001). As a result,  $D_C$  and  $D_A$  can be considered identical and we are left with only two sets of derivations: the set of derivations  $D$  and the set of converging derivations  $D_C$ . Proponents of the so-called crash-proof syntax framework claim that nothing more is needed; more specifically they claim that ‘no filters are imposed on the end products of derivations, and no global filters (e.g. comparison of derivations) assign status to derivations as a whole’; see Frampton and Gutmann (2002: 90) and the contributions in Putnam (2010) for extensive discussions of the viability of this claim. Chomsky (1995: ch. 4, 221), however, maintained the more traditional line of thinking by introducing bare output conditions, which are later referred to as interface conditions, which are ‘imposed from the outside’ by the performance systems that make use of the representations created by  $C_{HL}$ : the articulatory-perceptual and the conceptual-intentional system. Chomsky further claims that the interface conditions are involved in the displacement property of language and we will see below that he formulates these conditions in later work in the format of a filter on the output of  $C_{HL}$ ; cf. Chomsky (2001) and the discussion of (10/20) below.

We already noted that the early P&P models diverge from the linear model in Figure 1.1 in that the derivation of the PF- and LF-representations split at a certain point in the derivation in order to account by means of covert movement for the fact that there can be certain mismatches between linear order and semantic interpretation. Very early in the development of MP, proposals have been put forth to eliminate this property from the grammar. Groat and O’Neil (1996), for example, show that the copy theory of movement makes it possible to account for the discrepancies in PF and LF-representations by assuming that phonology can spell out either the lower or the higher copy in a movement chain; see also Bobaljik (2002). Chomsky (1995: ch. 4) argues that economy considerations can also account for these mismatches when we assume that it is more economical to move a syntactic category without its phonological features; pied piping of the phonological features is possible only when there are independent reasons to do so. The most recent development is the introduction of Agree (feature valuing at a distance) in the *Minimalist Inquiry* framework, which has made movement totally superfluous from a computational point of view. These proposals have in common that they make it possible to assume that the derivation of the LF- and PF-representations proceed in fully parallel fashion. The model of the Minimalist Inquiry framework, for example, is therefore as indicated in Figure 1.2.

Since Agree makes movement superfluous in the sense that it is no longer needed for feature checking, movement must be forced by other factors. More specifically, although movement must still be formally licensed by unvalued formal features, the question whether it actually applies depends on the interface conditions imposed by the conceptual-intentional (LF) or the articulatory-perceptual (PF) component on the output representations of  $C_{HL}$ . The intuition underlying this proposal is actu-



**Figure 1.2** The Minimalist Inquiry model (Chomsky 2000 and later).

ally much older than the Minimalist Inquiry framework. For example, it has been argued that the motivation for *wh*-movement is that a *wh*-phrase can only be interpreted if it heads an operator-variable chain; see for example Chomsky (1991: 440) and Rizzi (1996). What is new is that Chomsky (2001) claims that certain types of A-movement are also externally motivated. We will look at this in some detail in the remainder of this subsection.

According to MP, movement of a syntactic object S is subject to last resort: it must be triggered by some unvalued formal feature of a higher functional head H that can be checked or valued by a corresponding feature of S. In the earliest proposal it was assumed that these features of H come in two forms: weak and strong features. A strong feature on H must be checked before the projection of H is merged with some higher head; if checking does not take place, the derivation is canceled. A weak feature on H, on the other hand, cannot be checked before Spell-Out as a result of the economy condition Procrastinate. This proposal led to a very rigid system in which the question whether a certain movement does or does not apply is mechanically determined by the feature constellation of the functional head H. However, it is clear that movement may be sensitive to other factors as well. Consider the case of so-called object shift (OS) in the Icelandic examples in (6).

- |         |  |  |
|---------|--|--|
| (6) (a) | Jónke ypti ekki bókina.<br>Jón bought not the book       | <i>bókina</i> $\subset$ focus          |
| (b)     | Jón keypti bókina <sub>i</sub> ekki <i>t<sub>i</sub></i> | <i>bókina</i> $\subset$ presupposition |

The examples in (6) demonstrate that it is possible in Icelandic to move the direct object to the left, across the negative adverb *ekki*. This movement is not obligatory, however, but depends on the information structure of the clause: OS applies only when the object is part of the presupposition ('old' information) of the clause; it is excluded when it is part of the focus ('new' information) of the clause.

Let us provisionally assume that OS is triggered by the case feature on the light verb  $v^*$  (Vikner 1994; Chomsky 2001): if this case feature were strong, we would wrongly expect this movement to be obligatory; if it were weak, we would wrongly predict it to be impossible. In order to account for the apparent optionality of OS, we must therefore introduce additional means. One possibility would be to make the strength of the case feature sensitive to the information structure of the clause: only when the object is part of the presupposition of the clause does  $v^*$  have a strong case feature. Apart from being *ad hoc*, this option is not descriptively adequate since OS is never possible in complex tense constructions like (7): OS is excluded irrespective of the information structure of the clause, and (7a) is therefore ambiguous.





ment has some effect on the output representation. According to Chomsky this is only the case when the movement affects the semantic/pragmatic interpretation of the clause, or when it makes A'-movement possible (by placing the object at the phonological edge of the  $v^*P$ -phase). We will see shortly that this leads to a less rigid system in the sense that movement can be made sensitive to factors other than the feature constellation of the attracting head.

Chomsky claims that (10b) is also an invariant principle: in the terminology employed earlier, this claim expresses that an object occupying the position XP in (9) in the output representation must be construed as being part of the presupposition of the clause; see (8a). It is important to note that (10b) is only concerned with shifted objects, and leaves open the option that non-shifted objects are ambiguously interpreted as being part of either the focus or the presupposition of the clause. This is needed in order to allow the non-shifted objects in Icelandic examples like (7a) to be interpreted as part of the presupposition of the clause, and, of course, also correctly predicts that objects in languages like English, which do not have OS of the Icelandic sort, can be part of either the focus or the presupposition of the clause.

Given that (10b) does not restrict the interpretation of non-shifted objects, we need something in addition to account for the fact that OS is obligatory in examples like (6b). This is where (10c) comes in. Let us first consider the notion of phonological border, which is defined as in (11); since Chomsky does not specify the notion of *phonological material*, we take it to refer to an abstract set of phonological features that will be spelled out in the PF-component.

- (11) XP is at the phonological border of  $v^*P$ , iff:
- (a) XP is a  $v^*P$ -internal position, and;
  - (b) XP is not c-commanded by  $v^*P$ -internal phonological material.

The main difference between the examples in (6) and (7) is that in the former the main verb has moved out of  $v^*P$  into T, whereas in (7) it has not and thus occupies a  $v^*P$ -internal position. Example (7a) is therefore correctly predicted to be ambiguous: since the  $v^*+V$  complex is  $v^*P$ -internal and c-commands the object, clause (10c) does not apply and the object can be interpreted either as part of the focus of the clause (INT') or as part of the presupposition of the clause (INT). Example (7b) is consequently blocked by (10a) because OS has no effect on the outcome as the object can also be assigned the interpretation INT in its base position in (7a). Therefore, in constructions like (7), the EPP-feature can only be assigned to  $v^*$  if it is needed to enable A'-movement. In (6), on the other hand, there is no  $v^*P$ -internal phonological material that c-commands the position Obj. Consequently, if the object occupies this position, (10c) states that it must be assigned INT'. Movement of the object into the XP-position in (9) therefore has an effect on the outcome by licensing the interpretation INT, and (10a) consequently allows assignment of an EPP-feature to  $v^*$ .

It is important to note that statement (10c) clearly functions as a filter in the sense of Chomsky and Lasnik (1977). First, it is clear that it cannot be considered a condition on the derivation: when we would apply it to the intermediate stage in (9), the desired distinction between (6) and (7) could not yet be made locally, because the verb and the subject are moved out of the  $v^*P$  only at a later stage in

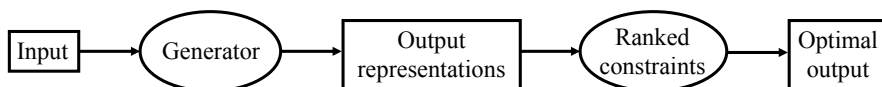
the derivation; Chomsky therefore assumes that it applies at the higher phase level (CP). Second, (10c) is a language-specific statement: Icelandic (and the continental Germanic languages) is subject to it, and therefore OS is forced in examples like (6b); the Romance languages, on the other hand, are not subject to it, so that (10a) blocks OS in comparable Romance examples. Thus, statement (10c) has the two characteristic properties of the PF-filters proposed in Chomsky and Lasnik (1977).

This subsection has shown that all grammars proposed during the P&P era have the global architecture of grammar indicated in Figure 1.1, although this was obscured in the early period by the assumption that derivations of the PF- and LF-representation diverge at some point in the derivation. It has been shown that by dropping this assumption Chomsky's recent Minimalist Inquiry framework fully conforms to the architecture in Figure 1.1; the grammar consists of a generative component that creates representations that are subsequently evaluated by a filter component. The filters place both semantic and phonological constraints on the output of  $C_{HL}$ , which reflects the fact that the representation(s) that pass these filters are subsequently fed to the articulatory-perceptual and the conceptual-intentional system where they undergo further computation in order to receive, respectively, a phonetic and a semantic interpretation.

## 2.2 Optimality theory

Optimality theory fits nicely to the global architecture of grammar in Figure 1.1, which is clear from the fact that it can be found in virtually all introductory texts on OT. It therefore also fits in the generative tradition as described in Section 2.1, but crucially differs from the P&P framework in that the evaluator is not taken to consist of universal principles and language-specific filters. The guiding intuition is instead that such principles and filters can be more adequately expressed by means of the ranking of a set of more primitive violable constraints; see Figure 1.3. We refer the reader to Pesetsky (1997; 1998) and Dekkers (1999) for early demonstrations of this.

Furthermore, OT adopts a holistic conception of language in the sense that the grammaticality of an expression E for some language L cannot be established by inspecting E alone, but is determined by comparing it to other expressions produced by the generator. This normally seems to go far beyond what is discussed under the term transderivationality in early minimalism; a derivation that is blocked by an economy constraint yields an ungrammatical expression in minimalism, whereas a loser in one OT-competition may still be the winner of another competition. It must be noted, however, that the latter is also a property of the set of statements in (10), which shows that Chomsky's most recent version of MP converges with this aspect of OT.



**Figure 1.3** Optimality theory.

Like the MP model in Figure 1.2, the OT model in Figure 1.3 entails two notions of well-formedness: one with respect to the generator and one with respect to the evaluator. The usual version of the generator in OT is, however, more liberal and unrestricted than  $C_{HL}$ , and allows for a comparatively large candidate set. The OT-evaluator, of course, differs from the one proposed in MP in that it uses ranked constraints instead of interface condition, but they do resemble them in that they often incorporate aspects of the interpretative systems; cf. Vogel (2004; this volume).

An important difference between OT and MP is that the former can also be seen as a meta-theory or a methodological guideline; this is clear from the fact that it may be applied to a wide variety of empirical domains: it can be equally well applied to phonology as to, for example, syntax, and it is certainly conceivable that it can also be successfully applied outside the domain of linguistics. When we restrict ourselves to a certain empirical domain, it may be that the differences between the different OT-approaches are so small that it is actually justified to speak of a more or less coherent theory. This might well be the case for OT-phonology, given that there seems to be considerable agreement among OT-phonologists on the nature of the input, the operations that can be performed by the generator, and the nature of the output. Furthermore, OT-phonologists do not only agree on the basic assumption that the evaluator consists of ranked violable constraints, but they also seem to share the belief that the postulated constraints are of just two types, the so-called faithfulness and markedness constraints. And, finally, there even seems to be some consensus about the individual constraints that are needed. Of course, there are also hotly debated issues, such as the question of whether the constraints are part of an innate, universally available set CON, or whether they are acquired on the basis of the primary linguistic data.

The situation in OT-syntax is entirely different: we are clearly not dealing with a generally accepted theory. Although Figure 1.3 is very specific about the nature of the evaluator, which has the defining property of consisting of ranked violable constraints, the nature of the generator is left open entirely; the generator can take the form of virtually any imaginable generative device, and, as a result, the generators of the current OT-approaches to syntax are based on different and often incompatible linguistic theories. Some more or less random examples are given in (12).

- (12) (a) Lexical-Functional Grammar: Bresnan (2000); Sells (2001)  
 (b) Early Principles-and-Parameters Theory: Grimshaw (1997); Pesetsky (1998)  
 (c) Minimalism: Dekkers (1999); Woolford (2007); Broekhuis (2008)  
 (d) Others: Müller (2000/2001); Vogel (2006)

Since the generators postulated by the proposals in (12) differ considerably and the generated candidate sets will therefore be constituted by candidates with entirely different properties, the postulated constraints will be quite different as well. As a result, we are dealing with OT-approaches that are as different as (or perhaps even more different than) the theories on which the generator is modeled. We will illustrate this below by comparing the OT-approaches proposed in Grimshaw (1997), Dekkers (1999) and Broekhuis (2008), which are all based on some version of the *principles-and-parameters* theory.

Grimshaw's (1997) proposal was originally written in the early 1990s and is based on the pre-minimalist *principles-and-parameters* framework. Among other things, this is clear from the fact that she tries to capture the directionality parameter, which was still generally assumed at that time, by means of two conflicting constraints HEAD LEFT and HEAD RIGHT (the head is leftmost/rightmost in its projection). In addition, she assumes the constraints SPECIFIER LEFT and SPECIFIER RIGHT (the specifier is leftmost/rightmost in its projection). Given that Grimshaw also assumes that the structures created by the generator conform the general X-bar-schema, the linearization of these structures follows from the language-specific ranking of these four constraints. Broekhuis (2008), which is based on the minimalist machinery proposed in Chomsky (2000) and later work, need not make use of Grimshaw's alignment constraints given that he adopts some version of Kayne's (1994) Linear Correspondence Axiom, according to which linear order is derived from the hierarchical relation between the constituent in the output representation. In his approach, linear order therefore follows from the language-specific ranking of a set of so-called EPP-constraints, which favor movement of a goal into its probe's minimal domain (in the sense of Chomsky 1995: ch. 3), and the economy constraint \*MOVE, which disfavors movement. For example, the 'strong' ranking  $EPP(case) \gg *MOVE$  requires movement of the probed noun phrase into the minimal domain of the unvalued case-features of  $v^*$  or the inflectional node I, whereas the 'weak' ranking  $*MOVE \gg EPP(case)$  requires that the probe remain in its original position. Such EPP-constraints, which are used to express the same intuition as Chomsky's Agree-based approach that Agree is normally sufficient for convergence, will find no place in OT-approaches that follow Groat and O'Neil (1996) in assuming that feature checking invariably triggers movement and that the linear order depends on the question whether it is the tail or the head of the resulting chain that is spelled out; such approaches will replace the EPP-constraints, for example, by Dekker's (1999) PARSE-F constraints, which favor pronunciation of moved constituents in the position of their formal features (the head of the chain), and reinterpret \*MOVE as a constraint that favors pronunciation of moved elements in their base position (the tail of the chain).

The previous paragraph has shown that properties of the proposed generator are immediately reflected in the nature of the postulated violable constraints of the OT-evaluator. The differences between the three OT-approaches discussed above are still relatively small due to the fact that the proposed generators all find their origin in the Chomskyan generative tradition, but it will be clear that the differences between these OT-approaches and OT-approaches that are based on other (generative) traditions may be much larger. For example, Broekhuis (2008) and Sells (2001) both develop an OT-analysis of Scandinavian object shift, but the two proposals differ at least as much as the minimalist and Lexical-Functional approaches that they are based on: whereas Broekhuis' analysis is built on the restrictions on movement of the clausal constituents, Sells' analysis is based on the restrictions on their phonological alignment.

Let us return to the guiding intuitions that connect all work in OT. The generator is an overgenerating system, which creates the candidate set from which the evaluator selects the optimal candidate(s) for a certain language L. The candidate set is

generally assumed to be infinite and to contain many candidates that will never surface because they are harmonically bound by some other candidate (where A is harmonically bound by B if A violates at least one constraint on top of the constraints violated by B). Furthermore, the focus of attention is on the evaluator, which consists of a set of constraints with the properties in (13a–c), which we will more extensively discuss below.

- (13) The optimality theoretic evaluator contains constraints that:
- (a) are taken from a universal set of constraints CON;
  - (b) are violable; and
  - (c) have a language-specific ranking.

The constraints crucially differ from the language-specific filters assumed in the *principles-and-parameters* theories in that they are generally assumed to be universal, that is, to be part of a universal set of constraints CON. These constraints can express language-specific properties by virtue of their properties in (13b) and (13c): languages may differ in the ranking of the universal constraints, and thereby select different candidates as optimal as a result of the fact that violation of a lower ranked constraint is tolerated in order to satisfy a higher ranked constraint. The way the OT-evaluator works can readily be demonstrated by means of Pesetsky's (1997; 1998) analysis of relative clauses. This will also give us the opportunity to show how the OT-evaluator differs from the filters assumed in the *principles-and-parameters* approaches. Consider again the relative clauses in (14) and (15), which were accounted for in *Filters and Control* by an appeal to the Doubly Filled COMP Filter and the recoverability condition on deletion.

- (14) (a) the man [<sub>COMP</sub> who ~~that~~ I know  $t_{\text{who}}$ ]  
 (b) the man [<sub>COMP</sub> ~~who~~ that I know  $t_{\text{who}}$ ]  
 (c) the man [<sub>COMP</sub> ~~who that~~ I know  $t_{\text{who}}$ ]  
 (d) \*the man [<sub>COMP</sub> who that] I know  $t_{\text{who}}$
- (15) (a) the book [about which ~~that~~ he spoke  $t_{\text{about which}}$ ]  
 (b) \*the book [~~about which~~ that he spoke  $t_{\text{about which}}$ ]  
 (c) \*the book [~~about which that~~ he spoke  $t_{\text{about which}}$ ]  
 (d) \*the book [about which that that he spoke  $t_{\text{about which}}$ ]

When we contrast these examples with the French relative clauses in (16) and (17), we see that English and French differ in that the former allows a wider variety of constructions with a bare relative pronoun than the latter. However, when the relative pronoun is embedded in a PP (or an NP), the two languages behave the same.

- (16) (a) \*l'homme [qui ~~que~~ je connais  $t_i$ ]  
 (b) l'homme [~~qui~~ que je connais  $t_i$ ]  
 (c) \*l'homme [~~qui que~~ je connais  $t_i$ ]  
 (d) \*l'homme [qui que je connais  $t_i$ ]

- (17) (a) l'homme [avec qui<sub>i</sub> que j'ai dansé t<sub>i</sub>]  
 (b) \*l'homme [avec qui<sub>i</sub> que j'ai dansé t<sub>i</sub>]  
 (c) \*l'homme [avec qui<sub>i</sub> que j'ai dansé t<sub>i</sub>]  
 (d) \*l'homme [avec qui<sub>i</sub> que j'ai dansé t<sub>i</sub>]

In order to account for the data in (14) to (17), Pesetsky proposed the constraints in (18), which we slightly simplify here for reasons of exposition. Constraint (18a) is simply the recoverability condition on deletion from Chomsky and Lasnik (1977), constraint (18b) is a constraint that expresses that embedded clauses tend to be introduced by a complementizer, and (18c) is a constraint that expresses that function words (like complementizers) tend to be left unpronounced.

- (18) (a) RECOVERABILITY (REC): a syntactic unit with semantic content must be pronounced unless it has a sufficiently local antecedent.  
 (b) LEFT EDGE (CP): the first leftmost pronounced word in an embedded CP must be the complementizer.  
 (c) TELEGRAPH (TEL): do not pronounce function words.

The ranking of these constraints will determine the optimal output. In order to see this, it is important to note that LE(CP) in (18b) and TEL in (18c) are in conflict with each other: the first wants the complementizer to be pronounced, whereas the latter wants it to be deleted. Such conflicts make it possible to account for variation between languages: when we rank these constraints differently, we get languages with different properties. When we assume that LE(CP) outranks TEL, we get a language in which embedded declarative clauses must be introduced by a complementizer. When we assume that TEL outranks LE(CP), we get a language in which embedded declarative clauses are not introduced by a complementizer. When we assume that the two constraints are in a tie (ranked equally high), we get a language in which embedded declarative clauses are optionally introduced by a complementizer. The evaluations can be made visible by means of tableaux. Tableau T1 gives the evaluation of embedded declarative clauses with and without a pronounced complementizer in a language with the ranking LE(CP) >> TEL.

**T1** No complementizer deletion in embedded declarative clauses

	LE(CP)	TEL
... [ complementizer ...] $\wp$		*
... [ <del>complementizer</del> ...]	*!	

The two asterisks indicate that the constraint in the header of their column is violated. The first candidate, with a pronounced complementizer, violates TEL but this is tolerated because it enables us to satisfy the higher ranked constraint LE(CP). The second candidate, with a deleted complementizer, violates LE(CP), and this is fatal (which is indicated by an exclamation mark) because the first candidate does not violate this constraint. The first candidate is therefore optimal, which is indi-

cated by means of the pointed finger: ☞. The shading of the cells indicates that these cells do not play a role in the evaluation; this convention is mainly for convenience, because it makes it easier to read the tableaux.

Now consider the evaluation of the same candidates in a language with the ranking  $TEL \gg LE(CP)$ , given in T2. Since  $TEL$  is now ranked higher than  $LE(CP)$ , violation of the former is fatal, so that deletion of the complementizer becomes obligatory.

### T2 Obligatory complementizer deletion in embedded declarative clauses

	TEL	LE(CP)
... [ complementizer ...]	*!	
... [ <del>complementizer</del> ...] ☞		*

Tableau T3 gives the evaluation of a language in which the two constraints are in a tie  $LE(CP) \triangleleft TEL$ , which is indicated in the tableau by means of a dashed line. Under this ranking, the rankings  $LE(CP) \gg TEL$  and  $TEL \gg LE(CP)$  are in a sense simultaneously active. Therefore we have to read the tie in both directions: when we read the tie from left to right, the violation of  $LE(CP)$  is fatal (which is indicated by  $>$ ), and the first candidate is optimal; when we read the tableau from right to left, the violation of  $TEL$  is fatal (which is indicated by  $<$ ), and the second candidate is optimal. This predicts that deletion of the complementizer is optional in this case.

### T3 Optional complementizer deletion in embedded declarative clauses

	LE(CP)	TEL
... [ complementizer ...] ☞		$<*$
... [ <del>complementizer</del> ...] ☞	$*>$	

Let us now return to the difference between English and French with respect to the pronunciation of relative clauses. It is clear that English has the tied ranking  $LE(CP) \triangleleft TEL$ , given that the complementizer is normally optional in embedded declarative clauses. In French, on the other hand, it is clear that  $LE(CP)$  outranks  $TEL$  given that the complementizer is obligatory in embedded declarative clauses. Pesetsky (1997) has shown that this also accounts for the differences between the English and French examples in (14) and (16), in which a bare relative pronoun is preposed. Assume that in both languages the constraint *RECOVERABILITY* outranks the constraints  $TEL$  and  $LE(CP)$ ; the ranking of the constraints in (18) are then as given in (19).

- (19) (a) French:  $REC \gg LE(CP) \gg TEL$   
 (b) English:  $REC \gg LE(CP) \triangleleft TEL$



The evaluation of the French examples in (16) proceeds as in T4. Since the relative pronoun has a local antecedent it is recoverable after deletion, so that all candidates satisfy REC. The second candidate is the optimal candidate because it is the only one that does not violate LE(CP); the fact that this candidate violates the lower-ranked constraint TEL is tolerated since this in fact enables the satisfaction of the higher-ranked constraint LE(CP).

#### T4 Relative clauses with preposed relative pronoun

French	REC	LE(CP)	TEL
l'homme [qui <sub>i</sub> <del>que</del> je connais t <sub>i</sub> ]		*!	
l'homme [ <del>qui</del> <sub>i</sub> que je connais t <sub>i</sub> ]	☞		*
l'homme [qui <sub>i</sub> <del>que</del> je connais t <sub>i</sub> ]		*!	
l'homme [qui <sub>i</sub> que je connais t <sub>i</sub> ]		*!	*

The evaluation of the English examples is slightly more complex than that of French due to the fact that LE(CP) and TEL are in a tie: we are therefore dealing with two rankings at the same time: REC >> LE(CP) >> TEL and REC >> TEL >> LE(CP). The first ranking is actually the one we also find in French, and we have seen that this results in selection of the second candidate as optimal. Under the second ranking, violation of TEL is fatal, so that the first and third are selected as optimal. As a result, three out of the four candidates are grammatical in English.

#### T5 Relative clauses with preposed relative pronoun

English	REC	LE(CP)	TEL
the man [who <sub>i</sub> <del>that</del> I know t <sub>i</sub> ]	☞	*>	
the man [ <del>who</del> <sub>i</sub> that I know t <sub>i</sub> ]	☞		<*
the man [ <del>who</del> <sub>i</sub> <del>that</del> I know t <sub>i</sub> ]	☞	*>	
the man [who <sub>i</sub> that I know t <sub>i</sub> ]		*>	<*

Next consider the evaluation of the French examples in (17), in which a PP containing a relative pronoun is preposed. Since the preposition is not locally recoverable, deletion of it leads to a violation of the highest-ranked constraint REC: this excludes the second and the third candidate. Since the two remaining candidates both violate LE(CP), the lowest ranked constraint TEL gets the final say by excluding the fourth candidate. Note that this shows that the ranking LE(CP) >> TEL does not mean that the complementizer is always realized, but that this may depend on other factors; when the complementizer is preceded by some element that must be realized, TEL forces the complementizer to delete.

## T6 Relative clauses with preposed PP

French	REC	LE(CP)	TEL
l'homme [avec qui <sub>i</sub> <del>que</del> j'ai dansé t <sub>i</sub> ] $\mathcal{P}$		*	
l'homme [avec qui <sub>i</sub> que j'ai dansé t <sub>i</sub> ]	*!		*
l'homme [avec qui <sub>i</sub> <del>que</del> j'ai dansé t <sub>i</sub> ]	*!	*	
l'homme [avec qui <sub>i</sub> que j'ai dansé t <sub>i</sub> ]		*	*!

For the English examples in (15) we get the same result as in French: both the second and the third candidate are excluded by REC, and the fourth candidate is excluded because it is harmonically bound by the first candidate: it has a fatal violation of TEL irrespective of the question whether we read the tie from left to right or from right to left.

## T7 Relative clauses with preposed PP

English	REC	LE(CP)	TEL
the book [about which <sub>i</sub> <del>that</del> he spoke t <sub>i</sub> ] $\mathcal{P}$		*	
the book [about which <sub>i</sub> that he spoke t <sub>i</sub> ]	*!		*
the book [about which <sub>i</sub> <del>that</del> he spoke t <sub>i</sub> ]	*!	*	
the book [about which <sub>i</sub> that he spoke t <sub>i</sub> ]		*	*!

## 2.3 Conclusion

This section has argued that the global architecture of grammar is as given in Figure 1.1, and that the several proposals made within the P&P approach do not differ in this respect from OT. The two frameworks are similar in assuming that we are dealing both with derivations and with evaluations: a generator creates a potentially multi-membered set of expressions S, and an evaluator determines which expressions from S are grammatical in a given language L. The OT view on the evaluator seems to be of a more optimistic nature than that of the P&P approaches. The latter consider the evaluator as a more or less random collection of language-specific filters on the output of narrow syntax. Pesetsky's work has shown, however, that at least some of the filters proposed by Chomsky and Lasnik (1977) can be decomposed into more atomic OT constraints, and Dekkers (1999) has shown this for a number of other constraints/principles like the ECP. Furthermore, since the OT constraints are claimed to be universal, they make precise predictions about the range of language variation that is allowed: Pesetsky, for example, has shown that his proposal is able to account for the differences between English and French relative clause constructions, and Broekhuis and Dekkers (2000) and Dekkers (1999) have shown that his proposal can be readily extended to relative constructions in Dutch.

### 3. Where MP and OT do differ: derivations and valuations

The previous section has shown that MP and OT assume the same global architecture of grammar, but there are, of course, also a number of obvious differences. This subsection will argue, however, that these do not have a principled linguistic motivation, but are the result of a more or less accidental difference in focus of attention between the two approaches: MP is mainly concerned with the universal (derivational) aspects of grammar whereas OT-syntax rather focuses on more language-specific aspects of grammar. This focus of interest is also reflected in the research strategies that the two approaches employ: research in MP tends to attribute to the generator  $C_{HL}$  as many properties of languages as possible, whereas Research in OT tends to appeal to the evaluator instead. It is therefore not surprising that the empirical successes of the two approaches also lie in different areas: MP is especially well equipped to account for the universal properties of languages, but there is no generally accepted view on the way we should account for, or even approach, the many ways in which languages may differ from each other; OT, on the other hand, precisely provides such a general theory of language variation, but since there is no generally accepted theory of the generator, current OT-syntax fails to account for the ‘truly’ universal properties of languages. These differences between MP and OT will be discussed more extensively below.

#### 3.1 Universal properties of language (the generator)

Both MP and OT-syntax hold the generator responsible for the invariant properties of language: the generator determines what representations are contained in the output, and hence can take part in the evaluation. The two frameworks differ, however, with respect to the extent that the generator is developed, or invoked in the analysis of the linguistic data.

The investigation of the generator ( $C_{HL}$ ) is considered MP’s core business. It has resulted in a sophisticated, restrictive theory on the nature of the generator. It is assumed that  $C_{HL}$  is constituted by a small set of operations that are subject to inviolable conditions that are relatively well understood. Perhaps  $C_{HL}$  can be reduced to a single merge operation, which has two incarnations, external and internal merge. As a result of this, the output of  $C_{HL}$  is also highly restricted; although it can be a non-singleton set, the differences between the members of this set are very limited in nature, and perhaps only involve the number of movements that occurred; cf. the discussion of Icelandic OS in section 2.1. It seems that analyses that do not invoke filtering devices are valued higher in MP than those that do, and, as a result, research tends to focus on those phenomena that can be successfully approached by means of a derivational account, with a concomitant reduction of the empirical scope of the theory, as is clear from the fact already mentioned earlier that Chomsky (1995: §4.7.3) suggests that ‘rearrangement’ phenomena like extraposition, right-node raising, VP-adjunction and scrambling are not part of narrow syntax. It is generally admitted in OT-syntax that the generator is the locus of the ‘truly’ universal properties of language: Grimshaw

(1997), for example, assumes that the structures formed by the generator conform to some version of X-bar-theory, Pesetsky (1998) and Anderson (2000) adopt some version of generative grammar as the generator, and Bresnan (2000) and Sells (2001) argue in favor of some version of Lexical Functional Grammar. The nature of the generator is, however, not a prominent subject of research; it is rather exceptional for an OT-researcher to account for some phenomenon by taking recourse to the generator given that most research in OT-syntax focuses on the variation that can be found rather than on the universal properties of languages.

Despite the differences in theoretical background (P&P, LFG, etc.), it seems that the view on the generator of many (if not most) OT-syntacticians crucially differs from that of MP-researchers, which becomes especially apparent when we consider the differences in the view on the output of the generator. We have already seen that, although MP allows for non-singleton output sets, MP-researchers generally take it for granted that this set is very small and that differences between the members of this set are limited in type, perhaps confined to differences in movement. In OT, on the other hand, it is generally maintained that the output of the generator can in principle be infinitely large, and that the members of the set may differ in a wide variety of ways. This suggests that the generator is generally taken to contain a larger set of operations in OT than is assumed in MP, and that these operations are probably confined in a less strict manner than the operations assumed in MP.

As a result of this different view on the generator, MP and OT tend to provide entirely different explanations for similar phenomena, the former taking recourse mainly to properties of the generator and the latter to those of the evaluator. This state of affairs seems to strengthen the widely accepted view that we are dealing with two competing and essentially incompatible frameworks. However, it can also be assessed differently, and more positively. Since it is not *a priori* given whether a certain phenomenon should be accounted for by appealing to the generator or the evaluator, it is important to develop alternative analyses that can subsequently be compared and evaluated; the fact that in some domains competing MP- and OT-analyses are available therefore does not mean in itself that we are dealing with competing or conflicting theories.

### 3.2 Variation (the evaluator)

One of the main concerns of both MP and OT is *cross-linguistic variation*. However, the way they approach this problem is entirely different – at least, at first sight. Let us start with discussing the way MP approaches the issue. Language variation is assumed to arise as a result of additional constraints on the application of the otherwise universal generator ( $C_{HL}$ ). The generator can basically perform two operations: external and internal merge. Let us provisionally adopt the standard assumption in MP that external merge is indispensable given that it is needed in order to assemble lexical items into semantically interpretable structures, for example, by the saturating the thematic roles of a given lexical head. Despite the fact that internal merge (movement) may have certain semantic implications, it is not essential in the

creation of semantically interpretable structures, so that we expect to find language variation in this domain. Note that since MP is mainly concerned with narrow syntax it mainly studies differences between languages that are somehow related to movement: variation in other domains is attributed to other modules (like PF), and is generally not discussed any further.

In early MP, the locus of variation between languages is solely attributed to the lexicon. Differences in the displacement property of languages are due to differences in the ‘strength’ property of the morpho-syntactic features that trigger movement: strong features trigger overt movement, whereas the weak features allow covert movement (which is favored by Procrastinate). In the more recent Agree-based theories, which reject the idea of covert movement, the core idea is preserved by assuming that movement only takes place if a functional head F contains an EPP-feature, which requires that the specifier of F be present. Under this view, the task of the language learner is to determine whether the functional head F has a weak or strong feature, or, alternatively, whether it has an EPP-feature, and to store this information in the lexicon.

The scope of OT goes much beyond the displacement property of languages: in principle, all (phonological, syntactic, semantic, pragmatic, etc.) properties can be fruitfully investigated, as long as one can plausibly postulate constraints bearing on the phenomenon in question. As we have already seen variation between languages is attributed to the evaluator, more specifically to the differences in ranking of the otherwise universal constraints. Under this view, the task of the language learner is to determine the constraint ranking (as well as the lexicon) of the language.

The discussion above seems to reveal another important difference between MP and OT: in the former, cross-linguistic variation is solely due to differences in lexical specifications, whereas in the latter it is rather due to the ranking of the universal constraints. This is indeed the case when we compare early MP with OT-syntax, but it no longer holds when we compare the most recent Minimalist Inquiry framework and OT-syntax. The early MP thesis that the sole locus of cross-linguistic variation is the lexicon runs into severe problems when we consider *variation within a single language*, because it predicts that languages cannot have ‘optional’ movement, that is, movements that occur only under well-defined semantic or phonological conditions. One example of this type of movement is Icelandic OS (already discussed in section 2.1), which can only apply when the object is part of the presupposition of the clause (cf. (6)), and when it does not cross the verb (cf. (7)) or other v\*P-internal material. This kind of optionality cannot arise under the early MP thesis because the postulation of feature strength or an EPP-feature gives rise to a very rigid system: when a feature is strong/an EPP-feature is present, movement must apply; when a feature is weak/an EPP-feature is not present, movement is blocked by Procrastinate.

We have already seen that this problem has led Chomky (2001) to assume that the EPP-feature, which forces movement, is optionally present. In order to avoid circularity, the choice must be made sensitive to external factors like the semantic and phonological conditions imposed on the pertinent movement, and this is precisely what Chomsky did in his account of OS in Icelandic in (10), repeated below as (20): as we have seen, the language-specific statement in (20c), in tandem with the

universal principles in (20a,b), precisely derives the circumstances under which Icelandic OS applies.

- (20) (a)  $v^*$  is assigned an EPP-feature only if that has an effect on outcome.  
 (b) The EPP position is assigned INT.  
 (c) At the phonological border of  $v^*P$ , XP is assigned INT'.

Chomsky (2001: 36) presents clause (20c) as a parameter that distinguishes OS from non-OS languages. French, for example, has verb movement to I, but nevertheless OS does not apply. This can be accounted for by assuming that (20c) does not hold for French. As a result, the interpretation INT can be assigned to the object when it is at the phonological border of  $v^*P$ ; as a result, movement of the object to the EPP-position is not needed and assignment of an EPP-feature to  $v^*$  is consequently blocked by (20a).

It seems, however, that (20c) is unlike the parameters of the earlier P&P framework in that it is not binary, because it is not the case that languages can be straightforwardly divided between OS and non-OS languages. This will become clear when we consider the Danish examples in (21) and (22), taken from Vikner (1994: 502); The examples in (21) show that Danish differs from Icelandic in that it does not have OS of non-pronominal DPs, whereas the examples in (22) show that it does have OS of weak pronouns.

- (21) (a) Hvorfor læste studenterne ikke artiklen?  
           why      read  the students not  the article  
 (b) \*Hvorfor læste studentene artiklen<sub>i</sub> ikke  $t_i$ ?
- (22) (a) Hvorfor læste studenterne den<sub>i</sub> ikke  $t_i$ ?  
           why      read  the students it     not  
 (b) \*Hvorfor læste studenterne ikke den?

This can be accounted for by assuming that clause (20c) must be further refined as in (20c'). This clause correctly expresses: (a) that non-pronominal DPs that are part of the presupposition of the clause (= INT) must undergo OS in Icelandic, but not in Danish or the Romance languages; and (b) that definite pronouns (which are assigned INT by definition) must undergo OS in Icelandic and Danish but not in the Romance languages.

- (20) (c') At the phonological border of  $v^*P$ , XP is assigned INT'  
 (i) XP = DP (Icelandic)  
 (ii) XP = definite pronoun (Danish)  
 (iii) XP =  $\emptyset$  (Romance)

What we want to stress here is that the adoption of language specific statements like (20c) or (20c') is a radical break with the early MP thesis that the sole locus of cross-linguistic variation is the lexicon. Since these statements essentially function as language-specific filters on the output of  $C_{HL}$ , linguistic variation should

be attributed to the evaluator in the model in Figure 1.2, and not to the lexicon. In fact, it seems that Chomsky's proposal makes it possible to eliminate the EPP-features entirely: when we assume that movement is subject to Last Resort but applies optionally, we could simply replace clause (20a) by the claim that movement is possible only if it has an effect on the outcome. This would make it possible to attribute cross-linguistic language variation *entirely* to the evaluator, just like in OT. In (23) we attempt to rephrase Chomsky's proposal such that reference to the notion of EPP-feature becomes superfluous.

- (23) (a) Movement is possible only if it has an effect on outcome.  
 (b) The derived object position is assigned INT.  
 (c) At the phonological border of  $v^*P$ , XP is assigned INT'.  
 (i) XP = DP (Icelandic)  
 (ii) XP = definite pronoun (Danish)  
 (iii) XP =  $\emptyset$  (Romance)

### 3.3 Conclusion

Since we have seen that MP and OT assume more or less the same global organization of grammar, we may conclude that the differences in the research strategies of MP and OT are somewhat accidental: as far as we can see, there are no theory-internal reasons for these frameworks to limit their investigation to respectively the generator or the evaluator. The fact that MP and OT occasionally provide alternative analyses for similar data as a result of these differences in research strategy does not follow from insurmountable theoretical differences between the two frameworks either, but simply reflects the fact that it is not *a priori* given whether a certain phenomenon belongs to the computational system or to the filter component of core grammar. Early MP and OT-syntax do seem to adopt conflicting views on the nature of variation between languages: the former adopts the thesis that language variation can be reduced to differences in the feature specifications of the lexical elements (feature strength/EPP-features), whereas the latter assumes that language variation is due to the evaluator, that is, to differences in constraint rankings. In Chomsky's current *Minimalist Inquiry* framework, however, the early MP thesis has been dropped: language variation is (also) attributed to parameters like (23c), which essentially function as language-specific filters on the output of  $C_{HL}$ . Current MP and OT therefore both attribute language variation to the evaluator, and the main difference between MP and OT boils down to the question whether the evaluator appeals to output filters or to ranked constraints.

## 4. Organization of the book

The discussion above has shown that MP and OT-syntax are actually much more alike than is generally assumed or one would think at first sight. The least one can

say is that they assume a similar global architecture of the grammar, and thus face a number of similar questions, like:

- (24) (a) What are the defining properties of the generator?  
 (b) What are the defining properties of the evaluator?  
 (c) What is the division of labor between the generator and the evaluator?  
 (d) What is the role of the articulatory-perceptual and the conceptual-intentional system?

We therefore have to ask ourselves whether it is still justified to consider MP and OT different, divergent programs, or whether it is possible to combine the best results of these programs into a single theory of grammar. The studies that will follow all discuss these questions from different perspectives and thus provide a collection of possible answers to these questions. Some studies in fact go beyond this and show alternative ways in which the output of the derivational system can be filtered. This introduction will not summarize the individual studies or address their contents in detail; for this we refer to the abstracts and the studies themselves. However, we will conclude this introduction by giving a brief discussion of the organization of the book as a whole and the grouping of the individual studies.

#### 4.1 Part I: combining MP with an OT-evaluation

The first group of studies presents work within the so-called derivation-and-evaluation framework, which explicitly combines MP and OT into a hybrid system: more specifically, it is argued that some version of the computational system of human language  $C_{HL}$  from MP functions as a generator that creates a restricted set of potentially well-formed expressions, which are subsequently evaluated in an optimally theoretic fashion by means of limited number of violable constraints. The chapter by Hans Broekhuis introduces this framework and illustrates it by means of a topic that was also extensively discussed in this introduction: Object Shift. He shows that Chomsky's proposal can be readily rephrased in optimality-theoretic terms and that this has a wide range of empirical consequences. The chapter by Gema Chocano and Mike Putnam discusses a number of restrictions on the licensing of parasitic gaps that are problematic in a purely derivational framework like MP but which fall out quite naturally under the hybrid approach by adding a single constraint to the inventory proposed by Broekhuis. Martin Salzmann discusses the dialectal and intra-speaker variation that can be found with dative resumption in relative clauses, that is, the use of a clause internal dative pronoun instead of a trace when the antecedent of the relative clause corresponds to the dative argument of the verb. Salzmann argues that the range of variation shows that locality is an inviolable condition on movement, but that the standard version of MP is nevertheless ill-equipped to handle the attested variation; he therefore concludes that hybrid systems like the derivation-and-evaluation framework are optimal for expressing the correct generalizations.



## 4.2 Part II: local and global optimization

As we have seen above one of the conspicuous differences between MP and traditional OT is that the former focuses on the derivation (the generator) whereas the latter focuses on the output representations of the generator (the evaluator). The first two studies collected in Part II show that this is by no means a necessary difference between MP and OT and that it is readily possible to formulate derivational versions of OT.

The first example of such a derivational version of OT is *Harmonic Serialism* (HS), which is illustrated by John McCarthy and Kathryn Pruitt on the basis of stress assignment. According to HS, the output of the generator is evaluated in a step-wise fashion by means of a Gen-Eval loop: at each point in the derivation the output of the generator is evaluated by the evaluator, after which the optimal output is sent back to the generator. This results in a process of local optimization that continues until the input and the optimal output representations are identical; this is the point of convergence. McCarthy and Pruitt explicitly compare the derivation of metrical structure in HS to the structure building operations found in MP, and claim that ‘both theories seek to explain the derivation of complex structures by deriving them via repeated application of simple operations under the control of an optimizing grammar.’

The chapter by Fabian Heck and Gereon Müller seems to fit seamlessly in this view given that they argue that the application of the structure building operations of MP are constrained by a local optimization in a way that comes very close to HS. They distinguish structure-building features, which trigger Merge, and probe features, which trigger Agree, and assume that the order of saturation of these features is determined in a local, stepwise fashion. Empirical evidence in favor of local optimization is provided by cross-linguistic differences in case assignment to internal and external arguments (that is, the difference between nominative-accusative and ergative-absolutive languages), agreement patterns in German DPs with prenominal dative possessors, the use of the German expletive *es*, and VP-topicalization. Heck and Müller suggest that many analyses that involve larger domains can be rephrased in terms of local optimization, but admit that there may be certain analyses that may require larger optimization domains.

Although the contributions by McCarthy and Pruitt and Heck and Müller show that local optimization has clear advantages when it comes to the application of the operations of the generator, there are cases where global optimization seems more suitable. Heck and Müller, for example, predict a dichotomy between nominative-accusative and ergative-absolutive languages, but Ellen Woolford shows that the distinction is not always clear-cut given that in some languages the expression of ergative case depends on certain contextual properties of the construction as a whole. For example, in Hindi and Nepali ergative case is only used in, respectively, perfective and individual-level contexts; since this information is probably not available or accessible during the derivation it seems less likely that this can be accounted for by means of local optimization.

Another case that requires global optimization involves the order restrictions on Scandinavian object shift constrictions (see Section 3 above). This is again

underlined by the discussion of object shift in Scandinavian remnant topicalization constructions by Eva Engels and Sten Vikner, who show that the optimization domain must be at least as large as CP. Their contribution also contains a discussion of the cyclic linearization approach by Fox and Pesetsky (2005), and they show that this approach is less well equipped to account for the set of data they discuss than their OT-approach.

The studies in this section suggest that we may need to postulate both local and global optimization. This need not be frowned upon with suspicion as this actually reflects the current distinction found within MP between local condition on the operations of the computational system  $C_{HL}$  and the set of conditions imposed by the articulatory-perceptual and the conceptual-intentional systems on output representations. It simply shows that OT-evaluations may be pervasive in the grammar in the sense that both types of restrictions may be rephrasable in terms of violable OT-constraints.

### 4.3 Part III: optimal design, economy and last resort in OT

Chomsky has stressed at a number of occasions that the name minimalist program is less felicitous given that minimalist considerations are a defining part of any scientific enterprise; minimalist concerns are therefore expected to play an important role in the development of the more traditional OT-approaches as well. The studies collected in this part of the book are good examples of this.

Vieri Samek-Lodovici argues that when it comes to cross-linguistic variation OT meets the requirement of optimal design better than the traditional versions of MP developed in the 1990s. He shows that under the assumption that no ranking of the universal, potentially conflicting (hence violable) constraints is inherently superior to any other, language variation is a predicted outcome of OT, whereas under the postulation of universal inviolable (hence non-conflicting) conditions language variation requires language-specific stipulations like Chomsky's (1993) postulation of weak/strong features on certain lexical items. Samek-Lodovici further points out that the existence of conflicting constraints is implied by the postulation of bare output conditions: since the sensory-motor and the conceptual-intentional system serve largely independent goals, there is no reason to exclude the possibility of conflicting interface constraints.

One way in which OT and minimalism can coexist and complement each other lies in OT's potential to model the interfaces between syntax, semantics and phonology/phonetics. A leading idea of minimalism is to reduce syntax to what is ultimately necessary to fulfill the needs of these interfaces. Therefore, the more elaborate the interfaces are constructed, the simpler the syntactic generator might be construable. Ralf Vogel concludes in his chapter that only very few of the specifically minimalist properties of the syntactic generator are necessary for OT's syntax generator, when one exploits OT as interface theory as much as possible. The conception of an OT grammar that he argues for organizes the mapping between semantic, syntactic and phonological/phonetic representations using violable and conflicting mapping constraints. The

core concept of OT that is relevant here is faithfulness, formulated in a corresponding theoretic way. Vogel shows that from this perspective the optimal syntactic representations are not necessarily the most economical ones, but rather those that correspond best to semantic and phonological/phonetic representations. He further argues that OT's notion of markedness is more adequate than MP's notion of economy. Unmarked syntactic structures can be seen as part of such maximally isomorphic mappings.

The notion of economy in MP is often linked to the notion of last resort: a certain operation can only be used when it is needed to arrive at a converging derivation. In early MP, for example, movement was claimed to be possible only when it serves to check and eliminate an uninterpretable feature of a certain sort. Jane Grimshaw criticizes this use of the notion given that we could simply remove the 'last resort' concept and state that movement is only possible when a certain feature is present (as is indeed assumed in the later versions of MP that assume optional EPP-features; see our earlier discussion of Chomsky's account of object shift). Grimshaw further claims that the notion of last resort can only receive a coherent interpretation in theories of optimization with constraint interaction. In fact, it is claimed that the notion is in fact entailed by such theories and that the use of *any* grammatical device is the result of last resort: it is the best that can be done in a particular configuration given a certain constraint ranking. This is illustrated by means of the choice between V-to-T/C, *do*-support and free tense (tense not supported by a verb) in a variety of constructions and languages. The discussion shows that, contrary to popular belief, *do*-support is not language-specific but arises in different circumstances in different languages.

#### 4.4 Part IV: the role of the interpretative components

Standard MP assumes that the interpretative components (PF and LF) impose certain conditions on the output of the generator, and we have further seen that such interface conditions can be readily expressed in an OT-fashion by means of ranked, violable constraints. The two studies in this part of the book propose alternative ways in which the interpretative component may affect the output of the generator.

Like Samek-Lodovici, Hedde Zeijlstra argues that the conditions imposed by the sensory-motor and the conceptual-intentional system on the output representations of the generator are necessarily in conflict and therefore (at least partly) violable. This gives rise to a tension that is solved by different languages in different ways, with language variation as a result. Zeijlstra postulates an inviolable principle of Full Legibility, which requires that all elements in the output be legible at the level of LF and PF, but which differs from Chomsky's (1995) Full Interpretation in that it allows legible but uninterpretable elements to be present; as a result, Full Legibility can be satisfied in more than one way. However, given that uninterpretable elements do not facilitate legibility, their number should be reduced as much as possible, and Full Legibility thus invokes simplicity measures that disfavor the presence of such elements. Since the set of legible elements differs for the level of representation (PF or LF) we are dealing with, the simplicity measures imposed on the output of the computational system may be in conflict: reduction of uninterpretable

elements at LF may result in an increase of uninterpretable elements at PF, and vice versa. Zeijlstra claims that this may result in more than one ‘optimal solution’ and that languages may select different solutions as the grammatical option. He claims that the actual choice is determined by the ancestry/acquisition of the language in question: the simplicity measures select the simplest grammar *compatible with the target language* (which may also account for the fact that some languages seem to select a suboptimal solution). Zeijlstra thus agrees with traditional OT that the interpretative components impose conflicting violable constraints on the output representations of the generator without, however, using the OT-formalism of constraint ranking in his account of the selection of the grammatical candidates for a given language L.

Kleanthes Grohmann argues that the PF-component determines how the copies of movement are spelled out. The basic hypothesis is, however, that this is done by means of inviolable, universal conditions on the output representations: he divides the clause in three mutually exclusive domains (in which respectively the thematic, agreement and discourse information is expressed). He further shows that when the moved element and its copy are within the same domain, the latter must be phonetically expressed; in other configurations an ‘elsewhere’ condition requires deletion of the copy. Earlier work has shown that this proposal may account for a wide range of phenomena, but Grohmann also shows that there is a small range of facts in which the copy is unexpectedly spelled out due to the intervention of other ‘independent constraints of the grammar’; cf. his (41b). It seems that such cases may be a good testing ground for some of the proposals discussed in this book.

## Acknowledgments

The majority of studies collected in this book were papers presented at the workshops on *Descriptive and Empirical Adequacy* held in Berlin (December 2005) and Leiden (February 2008). We like to thank everyone who made these events possible as well as the contributors to the present book. We especially thank our assistants Ann-Christin Broschinski, Anna Kutscher, and George Moore, who did an amazing job in the final preparations of the manuscripts. Ben Hermans helped us with editing the manuscript by John McCarthy and Kathryn Pruitt.

We hope that this book will encourage people who are currently exclusively working in MP or OT to broaden their vision by also considering the possibilities that the alternative framework offers, and that this will stimulate further fruitful debate among proponents of the two frameworks. The research of Hans Broekhuis for this book was partly funded by a so-called VIDJ-grant (2003-8) from the Netherlands Organization of Scientific Research (NWO).

## References

- Anderson, S. R. (2000) Towards an optimal account of second-position phenomena. In J. Dekkers, F. Van der Leeuw and J. Van de Weijer (eds) *Optimality Theory: Phonology, Syntax and Acquisition*, 302–33. Oxford: Oxford University Press.
- Bobaljik, J. D. (2002) A-chains at the PF-interface: Copies and ‘covert’ movement. *Natural Language & Linguistic Theory* 20: 197–267.
- Bresnan, J. (2000) Optimal syntax. In J. Dekkers, F. Van der Leeuw and J. Van de Weijer (eds) *Optimality Theory: Phonology, Syntax and Acquisition*, 334–85. Oxford: Oxford University Press.

- Broekhuis, H. (2008) *Derivations and Evaluations: Object Shift in the Germanic Languages*. Berlin/New York: Mouton de Gruyter.
- Broekhuis, H. and Dekkers, D. (2000) The minimalist program and optimality theory: Derivations and evaluations. In J. Dekkers, F. Van der Leeuw and J. Van de Weijer (eds) *Optimality Theory: Phonology, Syntax and Acquisition*, 386–422. Oxford/New York: Oxford University Press.
- Chomsky, N. (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1986) *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, N. (1991) Some notes on economy of derivation and representation. In R. Freidin (ed.) *Principles and Parameters in Comparative Syntax*, 417–54. Cambridge, MA: MIT Press.
- Chomsky, N. (1995) *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2000) Minimalist inquiries: The framework. In R. Martin, D. Michaels and J. Uriagereka (eds) *Step by Step. Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.
- Chomsky, N. (2001) Derivation by phase. In M. Kenstowicz (ed.) *Ken Hale. A Life in Language*, 1–52. Cambridge, MA: MIT Press.
- Chomsky, N. and Lasnik, H. (1977) Filters and control. *Linguistic Inquiry* 8: 425–504.
- Dekkers, J. (1999) *Derivations & Evaluations. On the Syntax of Subjects and Complementizers*, Doctoral dissertation. University of Amsterdam/HIL.
- Fox, D. and Pesetsky, D. (2005) Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31: 1–45.
- Frampton, J. and Gutmann, S. (2002) Crash-proof syntax. In S. D. Epstein and T. D. Seely (eds) *Derivations and Explanations in the Minimalist Program*, 90–105. Malden, MA and Oxford: Blackwell Publishing.
- Grimshaw, J. (1997) Projection, heads and optimality. *Linguistic Inquiry* 28: 373–422.
- Groat, E. and O’Neil, J. (1996) Spell-Out at the LF interface. In W. Abraham, S. D. Epstein and H. Thráinsson (eds) *Minimal Ideas. Syntactic Studies in the Minimalist Framework*, 113–39. Amsterdam and Philadelphia, PA: John Benjamins.
- Holmberg, A. (1999) Remarks on Holmberg’s generalization. *Studia Linguistica* 53: 1–39.
- Kayne, R. (1994) *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Müller, G. (2000) Shape conservation and remnant movement. In A. Hirotnani, N. Hall Coetzee and J.-Y. Kim (eds) *Proceedings of NELS 30*, 525–39. Amherst, MA: GLSA.
- Müller, G. (2001) Order preservation, parallel movement, and the emergence of the unmarked. In G. Legendre, J. Grimshaw and S. Vikner (eds) *Optimality-theoretic Syntax*, 113–42. Cambridge, MA and London: MIT Press/MITWPL.
- Pesetsky, D. (1997) Optimality theory and syntax: Movement and pronunciation. In D. Archangeli and T. Langendoen (eds) *Optimality Theory*, 134–70. Malden, MA and Oxford: Blackwell.
- Pesetsky, D. (1998) Some optimality principles of sentence pronunciation. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky (eds) *Is the Best Good Enough?*, 337–83. Cambridge, MA and London: MIT Press/MITWPL.
- Putnam, M. (2010) *Exploring Crash-proof Grammars*. Amsterdam: John Benjamins.
- Rizzi, L. (1996) Residual Verb Second and the Wh criterion. In A. Belletti and L. Rizzi (eds) *Parameters and Functional Heads. Essays in Comparative Syntax*, 63–90. Oxford and New York: Oxford University Press.
- Sells, P. (2001) *Structure Alignment and Optimality in Swedish*. Stanford, CA: CSLI Publications.
- Vikner, S. (1994) Scandinavian object shift and West Germanic scrambling. In N. Corver and H. van Riemsdijk (eds) *Studies on Scrambling. Movement and Non-movement Approaches to Free Word-Order Phenomena*, 487–517. Berlin and New York: Mouton de Gruyter.
- Vogel, R. (2004) Correspondence in OT syntax and minimal link effects. In G. Fanselow, A. Stepanov and R. Vogel (eds) *Minimality Effects in Syntax*, 401–42. Berlin: Mouton de Gruyter.
- Vogel, R. (2006) The simple generator. In H. Broekhuis and R. Vogel (eds) *Optimality Theory and Minimalism: A Possible Convergence? Linguistics in Potsdam* 25, 99–136. Potsdam: University of Potsdam: <http://www.ling.uni-potsdam.de/lip>.
- Woolford, E. (2007) Case locality: Pure domains and object shift. *Lingua* 117: 1591–616.