From the origins of Government and Binding to the current state of Minimalism

Artemis Alexiadou[1,2] & Terje Lohndal[3,4]

[1]Humboldt Universität zu Berlin
[2]Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS)
[3]NTNU Norwegian University of Science and Technology
[4]UiT The Arctic University of Norway

Abstract[i]
This chapter provides a review of the current Chomskyan approach to the study of human language, known as the Minimalist Program. It offers an overview of the central ideas that were central in shaping the program, in particular Government and Binding Theory. It presents an outline of what the essential ideas of the program are, focusing in particular on how the Minimalist Program is seen as a natural development of Government and Binding Theory. Then it discusses a few current developments within the program. In general, the chapter aims to provide an overview in broad strokes, and therefore it focuses on conceptual and theoretical issues rather than technical or empirical results.

1. Setting the scene

Generative grammar is an approach to the study of language which is explicit , mentalistic, and based on the claim that the ability to acquire language is innately specified.[ii] The approach is concerned with language as a psychologically real object, whose representations can be studied scientifically. In the words of Chomsky (1975, 160): 'Linguistics is simply that part of psychology that is concerned with one specific class of steady states, the cognitive structures that are employed in speaking and understanding'. Put differently, '[w]e have grammars in our heads' (Smith and Allott 2016, 128). We humans come pre-wired with the ability to create these grammars: We are born with a unique ability to acquire language.

Throughout its history, three questions have been at the center of this approach. They are given in (1).[iii]

(1) a. What constitutes knowledge of language?
b. How is knowledge of language acquired?
c. How is knowledge of language put to use? (Chomsky 1986)

The first question seeks to establish the basis for our linguistic ability. Linguists often speak of this in terms of 'knowledge', but this is not the kind of knowledge that many philosophers will have in mind. Chomsky (1982, 128) says the following:

> As I am using the term, knowledge may be unconscious and not accessible to consciousness. It may be 'implicit' or 'tacit'. No amount of introspection could tell us what we know, or cognize, or use certain rules or principles of grammar, or that use of language involves mental representations formed by these rules and principles. We have no privileged access to such rules and representations.[iv]

Furthermore, generative scholars are interested in how children are able to acquire these representations for each variety or language. Lastly, an important question is how we humans utilize these representations in language use.

Since its inception, generative grammar has made at least three fundamental contributions to our understanding of language: i) viewing grammars as formal/mathematical objects, ii) viewing linguistics as psychology and biology: studying the emergence and structure of the mental architecture underlying language, iii) Universal Grammar, the proposal that there is innate, mental structure which is specific to language and that enables children to acquire any language. These results have partly come about through the focus on descriptive adequacy, descriptions of the intrinsic competence of a speaker, and on explanatory adequacy, how a child acquires this intrinsic competence:[v]

> To the extent that a linguistic theory succeeds in selecting a descriptively adequate grammar on the basis of primary linguistic data, we can say that it meets the condition of *explanatory adequacy*. That is, to the extent, it offers an explanation for the intuition of the native speaker on the basis of an empirical hypothesis concerning the innate predisposition of the child to develop a certain kind of theory to deal with the evidence presented to him.' (Chomsky 1965, 25-26; his italics)

This enabled the study of grammars that humans have internalized, unlike, say, studying a finite corpus.

In this chapter, we will outline some of the recent history leading up to contemporary generative grammar. We will first provide some context for the emergence of Principles and Parameters, before we provide the basic gist of the Principles and Parameters approach. Then we introduce the first model that was proposed, namely Government and Binding. This is followed by a discussion of the second and to this date current model, the Minimalist Program, before we try to outline some of the current trends that shape the field of generative grammar. Lastly, we summarize and conclude the chapter.

2. Some context: The emerging idea of Principles and Parameters

A core aspect of generative grammar in its early days was a computational system in the human mind that contained phrase structure rules for building hierarchical structures and more complex operations that were able to modify these phrase structures. The latter were known as transformations, and transformations crucially operated on structures, not, say, sentences as in Harris (1951). This gave rise to the name 'Transformational Grammar' which is synonymous with generative grammar. Transformations made the theory much more powerful as it allowed an infinite number of grammars (cf. Lasnik 2000, 114; Lasnik & Lohndal 2013, 27-28), raising serious questions regarding learnability: How can a child select the correct target grammar from all available grammars? In this section, we will summarize some of the most important context leading up to the Principles and Parameters approach, which we will present in section 3. For reasons of space, the present section will have to set aside a lot of details, but see Freidin and Lasnik (2011) and Lasnik and Lohndal (2013) for a more detailed exposition.

The grammatical architecture proposed in Chomsky (1965) looks as in (2) (Chomsky 1965, 135-136, cf. also Lasnik and Lohndal 2013, 36).

(2)

Deep Structure	⇒	Semantic Interpretation
⇓		
Transformations (operating cyclically)		
⇓		
Surface Structure	⇒	Phonetic Interpretation (via the "sound-related"

levels of Morphophonemics, phonemics, and phonetics)

To give one example, we can consider simple questions. In Chomsky (1955, 1957), (3a) and (3b) had the same initial phrase structure (called 'phrase marker' at the time).

(3) a. Ellie will solve the problem.
b. Will Ellie solve the problem?

Transformations take the structure of (3a) and transforms it into the structure of (3b). The details are not important; readers can consult Lasnik (2000) for a lucid exposition. A remarkable success of this approach, as Lasnik (2005, 69) emphasizes, is that it enabled a unified analysis of (3) and (4).

(4) a. Ellie solved the problem.
b. Did Ellie solve the problem?

Native speakers can sense a relationship between (3b) and (4b), but prior to Chomsky's analysis, there was no account of this. In Chomsky (1965), the technicalities were different, but the intuitions were the same: A common underlying Deep Structure as the basis for both declaratives and interrogatives, and then transformations that altered the structure into a Surface Structure, followed by morphophonological operations that provide the accurate forms for Phonetic Interpretation.

Transformations are an essential and powerful part of this architecture. Because of this, work conducted in the late 1960's and 1970's suggested a range of constraints to limit the power of transformations and consequently the range of possible grammars. An example of this is the work by Ross (1967) which proposed constraints on long-distance dependencies (Ross labeled them 'islands'; see Boeckx 2013, den Dikken and Lahne 2013 and Müller this volume for overviews). Nevertheless, as Chomsky and Lasnik (1977) point out, the quest for descriptive adequacy led to a tremendously rich theory. This can be seen quite clearly in Peters and Ritchie (1973), whose explicit formalization contains a range of mechanisms that were proposed at the time, such as global rules and transderivational constraints. Let us look at these mechanisms briefly (building on the discussion in Lasnik and Lohndal 2013).

Lakoff (1970, 628) defines a global rule as a rule that states conditions on 'configurations of corresponding nodes in non-adjacent trees in a derivation'. In general, transformations have always been assumed to be Markovian, that is, that they involve one step at the time. However, global rules require a system that dramatically extends the power beyond Markovian properties. Ross (1969) famously provided an example of a global rule. In this paper, he extends results he obtained in Ross (1967) involving island constraints. One such island constraint is illustrated in (5), the Coordinate Structure Constraint, which prevents extraction from just one of the conjunctions. We have illustrated that by showing in (5) a copy of *who* in the position from which it has been deleted.

(5) *Irv and someone were dancing, but I don't know who Irv and ~~who~~ were dancing.

Notably, Ross (1969) showed that if the constraint isn't visible, it goes away. A way to make it disappear is to use ellipsis, as in (6).

(6) Irv and someone were dancing, but I don't know who.

In (6), the coordinate structure, the constituent that forms the island, has been elided and is not pronounced. That makes the example acceptable. More formally, Ross argued that for an island violation to occur, the constituent that forms the island needs to be present at Surface Structure. If a transformation deletes this constituent, the constraint no longer applies. This deletion became known as 'sluicing' (see van Craenenbroeck and Merchant 2013). To capture the contrast between (5) and (6), island constraints need to mention both the surface structure and the point in the derivation where the movement of the relevant constitutent (*who* in (5)) takes place, the coordinate structure in (5). That the constraint needs to mention both properties makes it a global rule.

As for transderivational constraints, such constraints depend on derivations different from the one that is being considered. Hankamer (1973) provides arguments in favor of such constraints. One example involves the phenomenon known as 'gapping' (see van Craenenroeck and Merchant 2013). Among others, he used the example in (7) (Hankamer 1973, 26-27).

(7) Max wanted Ted to persuade Alex to get lost, and Walt, Ira.

The question is how such a string is derived, that is, what is the correct derivation underlying (7)? Possible candidates could be (8a) or (8b).

(8) a. Max wanted Ted to persuade Alex to get lost,
*and Walt [wanted] Ira [to persuade Alex to get lost]
b. Max wanted Ted to persuade Alex to get lost,
*and Walt [wanted Ted to persuade] Ira [to get lost]

Hankamer argued that both options in (8) are out because (7) can also be derived from a different constituent structure which still derives the intended meaning, namely (9).

(9) Max wanted Ted to persuade Alex to get lost,
and [Max wanted] Walt [to persuade] Ira [to get lost]

When the bracketed constituents are deleted, (9) becomes (7). Given this, the constraint would not just have to make reference to alternative derivations created from the same Deep Structure, but also to alternative derivations created from *different* Deep Structures. That raises non-trivial questions concerning the expressive power of such a computational system, and consequently also its learnability.[vi]

Any extension of the class of possible grammars requires significant empirical justification. Chomsky & Lasnik (1977) argued that this justification had not been provided in approaches that extended the original framework in Chomsky (1955/1975, Peters and Ritchie (1973), and comparable work, cf. Dougherty (1973), Chomsky (1973), and Brame (1976). Because of that, Chomsky & Lasnik proposed a new framework which restricted the number of possible grammars significantly. This was seen as a crucial step towards being able to explain the acquisition of grammatical competence, a central goal ever since Chomsky (1965).

The new framework departed from earlier frameworks in some crucial ways, not at least in assuming that Universal Grammar is not an 'undifferentiated' system. That is, it was argued that core grammar has highly restricted options, since it consists of universal principles and a few parameters that account for variation. In addition to the core, there is the periphery, consisting of 'marked' phenomena, e.g., irregularities (i.e., irregular verbs) and exceptions more generally (e.g., English has prepositions, but also the marked exception *ago*

– which comes *after* its complement). In other words, the approach required something similar to a theory of markedness, with all its complications (see Haspelmath 2006 for a comprehensive discussion). As Chomsky and Lasnik (1977, 430) say:

> Systems that fall within core grammar constitute "the unmarked case"; we may think of them as optimal in terms of the evaluation metric. An actual language is determined by fixing the parameters of core grammar and then adding rules or rule conditions, using much richer resources, perhaps resources as rich as those contemplated in the earlier theories of [transformational grammar][vii]

Research was generally devoted to the core phenomena: 'A reasonable approach would be to focus attention on the core system, putting aside phenomena that result from historical accident, dialect mixture, personal idiosyncrasies, and the like' (Chomsky and Lasnik 1993, 510).

The name for constraints in Chomsky and Lasnik (1977) was 'filters'. In their paper, the hypothesis was that surface filters can capture effects of ordering, obligatoriness and contextual dependencies. Such surface filters would be universal; thus, we would not expect any variation between languages. This makes filters different from parameters. Furthermore, a third component was language-specific filters. For example, to capture the ill-formedness of (10a) in Standard English, the language-specific filter in (10b) was proposed.

(10) a. *We want for to win.
b. *[for-to]

This filter deems any *for-to* string illicit. Chomsky and Lasnik (1977, 442) claim that the rule in (10b) would be a 'dialect' filter, since it was assumed to involve 'a high degree of uncertainty and variation'. And, indeed, *for to* sequences are perfectly possible in for example Irish English dialects. In essence, then, a filter can either be outside of core grammar, like (10b), or part of core grammar, like the ban on stranding an affix (The Stranded Affix Filter, cf. Lasnik 1981).

Chomsky and Lasnik's (1977) paper prepared the ground for a major change in how to think about universality and variation. We turn to that in the next section.

3. Principles and Parameters: Solving Plato's problem

Chomsky (1981) is a fundamental contribution to the study of human language in its effort to develop a new theory of what is universal across languages and what is variable. The main change affected the notion of filters, which came to be replaced by parameters. Parameters were seen as providing the solution to two issues: How can we capture the observed variation across the world's languages, and how do humans know so much given the limited evidence that is available to us (Plato's problem). The idea was that the child only had to set the correct value, which mostly was thought to involve a choice between two options, much like a switchbox as James Higginbotham put it. The head parameter is a simple example of this: You look at whether the verb precedes or follows the object, which gives you the two main word orders across the world's languages: verb-object, or object-verb. As Chomsky pointed out:

> If these parameters are embedded in a theory of UG that is sufficiently rich in structure, then the languages that are determined by fixing their values one way or another will appear to be quite diverse (Chomsky 1981, 4).

Importantly, on this approach, the space of linguistic variation is part of UG.

An essential part of the framework was clustering, that is, parameters should represent clusters of properties. Chomsky (1981, 6) put it as follows: '[I]deally we hope to find that complexes of properties […] are reducible to a single parameter, fixed in one or another way'. A prominent example of this reasoning is provided by the Null Subject Parameter (Rizzi 1982; see also D'Alessandro 2015 for a review), which governs the realization of silent subjects, like in Spanish and many other languages. An example is provided in (11), where the subject does not have to be pronounced. This contrasts markedly with the same sentence in English, where the subject cannot be left out.

(11) (Voi) state leggendo un libro. (Italian)
you.PL are reading a book
'You are reading a book.' (D'Alessandro 2015)

Rizzi (1982) argued that being a null subject language is correlated with several other properties: i) the subject can be moved out of a finite embedded clause that is headed by an overt complementizer, (12), ii) they allow for subject inversion (Kayne 1980), (13), iv) these are languages with so-called rich agreement on the verb (Taraldsen 1980; Alexiadou & Anangostopoulou 1998), in which both referential and non-referential (so-called expletive) null subjects are licensed, unlike in non-null subject languages, (14).

(12) ¿Quién$_i$ dijiste que t_i salió temprano? (Spanish)
who say-PRET.2.SG THAT leave-PRET.3.SG early
'Who did you say that left early?' (Perlmutter 1971, 103)

(13) a. È arrivato Gianni. (Italian)
b. *Est arrivé Jean. (French)
c. *Has arrived John. (English) (Roberts 2007, 28)

(14) a. It rains frequently in April.
b. *Rains frequently in April.
c. *Ello/Lo llueve a menudo en abril. (Spanish)
it.STRONG/it.CLITIC rain.3SG frequently in April
d. Llueve a menudo en abril.
'It rains frequently in April.' (Judy and Rothman 2010, 200-201)

This clustering highlights another virtue of the clustering idea, namely that properties that are hard to observe for the child are linked to properties that are easy to observe. It is easy for a child to determine whether the language in question is a null subject language or not, but it is not easy to observe the relevant data that will tell her that extraction out of a finite embedded clause with an overt complementizer is licit or not, as in (12). Such data are not very frequent and this may create problems when it comes to extracting the right generalization from them when they are observed. If the clustering claim is true, the model really would have solved Plato's problem in abstraction: Learners do not have to compare whole grammars, they need to learn one parameter after the other. This would be a small number of choices, each of which would not involve many options. However, this only holds in abstraction because the model was never supplemented with an actual theory of how the child would learn the relevant parameter settings from the input.

We said 'if true' because a lot of work since has cast a lot of doubt on the particular clustering that Rizzi (1982) argued for (e.g., Haspelmath 2008, Jaeggli and Hyams 1988, Newmeyer 2005, Rothman and Iverson 2007; Rothman 2009a, Sheehan this volume). Baker (2008, 352) claims that '[h]istory has not been kind to the Pro-drop Parameter as originally

stated'. Since Chomsky (1981), the mainstream view on parameters has changed quite a bit, as Sheehan (this volume) illustrates and which we will return to briefly below.

The first part of Principles and Parameters was immensely productive, in particular from a cross-linguistic point of view. Hornstein (2013, 399) puts this well in the following quote:

> It is important to appreciate how fecund and productive this period of research [up to the 1990's] was. Arguably, we learned more new facts about more typologically diverse languages than ever in the history of the study of language. We learned a tremendous amount about how languages and grammars operate, what they have in common and how they differ.

This is an important result that is often underappreciated by Chomsky's critics.

The Principles and Parameters theory came to have two different instantiations or models (cf. Freidin and Vergnaud 2001, Hornstein, Nunes and Grohmann 2005, Lasnik and Uriagereka 2005, Boeckx and Uriagereka 2006, Lasnik & Lohndal 2010, 2010, Freidin 2012). The first one was called Government and Binding, and this was the model that was developed and used throughout the 1980's. In the beginning of the 1990's, Chomsky developed a new model, called the Minimalist Program. We will discuss the essentials of Government and Binding in the next section, and then in section 5, explain the transition to the Minimalist Program.

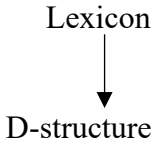
4. Government and Binding

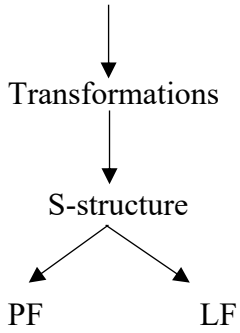
The main idea behind the model of grammar labeled Government and Binding (GB) was developed by Chomsky in several publications (1981, 1982, 1986). According to Chomsky (1981: 5ff.), the theory of grammar consists of a rule system and its sub-components, depicted in (15), and a system of principles and its subcomponents, shown in (16). The overall architecture is provided in (17).

(15) Rule system
- 1. lexicon
- 2. syntax
- a. categorial component
- b. transformational component
- 3. Phonetic Form
- 4. Logical Form

(16) Principle system
- i. bounding theory
- ii. theta-theory
- iii. binding theory
- iv. goverment theory
- v. case theory
- vi. control theory

(17) Y-model





In constructing a sentence (what is typically called the 'derivation'), the grammar machine starts out with lexical and functional items from the lexicon. The lexicon contains all idiosyncratic information about lexical items. In turn, these items feed the syntax by way of assembling a D-structure where argument structure properties are assigned, i.e., a transitive verb will be in a configuration with an object and a subject, and an intransitive verb with just a subject. Then transformations (called Move α, where 'α' stands for any constituent) map this structure into an S-structure (e.g., in the case of *wh*-movement as in *What did Jane read?*, which starts out with the structure *Jane read what*), which fairly closely resembles the surface order of words. At this point, the resulting structure is transferred to the two interpretive interfaces: Phonetic Form (sound) and Logical Form (meaning). Note also that the principal subsystems in effect apply wellformedness constraints on the other levels of representation. Chomsky (1981, 5-6) summarizes the organization of the grammar as follows.

> The lexicon specifies the abstract morpho-phonological structure of each lexical item and its syntactic features, including its categorial features and its contextual features. The rules of the categorial component meet some variety of X-bar theory. Systems 1.1 and 1. 2a constitute the base. Base rules generate D-structures (deep structures) through insertion of lexical items into structures generated by 1.2a, in accordance with their feature structure. These are mapped to S-structure by the rule Move-alpha, leaving traces coindexed with their antecedents; this rule constitutes the transformational component 1. 2b, and may also appear in the PF- and LF-components. Thus the syntax generates S-structures which are assigned PF- and LF-representations by components 3 and 4 of 1. Bounding theory poses locality conditions on certain processes and related items. The central notion of government theory is the relation between the head of a construction and categories dependent on it. Theta-Theory is concerned with the assignment of thematic roles such as agent-of-action, etc.... Binding theory is concerned with relations of anaphors, pronouns, names, and variables to possible antecedents. Case theory deals with assignment of abstract Case and its morphological realization. Control theory determines the potential for reference of the abstract pronominal element PRO.

Much more can be said about each of these components, and Haegeman (1994) provides an excellent and detailed exposition. Many other textbooks also exist, so in the interest of space, we will not provide a more detailed outline of the Government and Binding architecture. Some necessary details will be covered in the discussion of the Minimalist Program below, but rather than go through the entire system here, we want to highlight a more general aspect of the model.

The above picture demonstrates the division of the language module into distinct submodules. In addition, it is assumed that the language faculty is a module of the mind distinct from other modules such as for instance the visual system. Chomsky's notion of

modularity is very different from Fodorian modules (Chomsky 2018 compares his position to Fodor). For Fodor, human cognition consists of a "central system" and a number of "input systems" (Fodor 1983). The Fodorian modules are cognitive 'reflexes' – specialized for particular domains, which operate fast and mandatorily, are hard-wired for a particular area of the brain, their structure and function largely innately determined, and they are informationally encapsulated. Fodor's view on language is that it is an i) input system analogous to those devoted to the senses, ii) the central system, however, is unstructured and uninvestigable.[viii] By contrast, Chomsky assumes that language is a central system used in both input and output (in expressing and communicating thoughts), and there are probably several such central systems: a language faculty; a number 'sense', and possibly others (see e.g., Fodor 2000, Smith and Allott 2016, and Chomsky 2018 on similarities and differences between their views).

While the framework of Government and Binding led to very rich empirical generalizations and discoveries (see also the chapters by Baker, Müller, Sheehan in this volume), there are some issues with its architecture. In particular the proliferation of internal modules and the fact that it postulates four levels of representation can be seen as violating the principle that one should only postulate the 'barest essentials' necessary for a theory of grammar. In addition, a new question on the agenda became how to account for the biological or evolutionary origin of language: To what extent is our theory of human language compatible with what we know about language evolution and the biological scaffolding of language? We turn to a discussion of these issues in the next section.

5. The Minimalist Program: Untying the descriptive vs. explanatory knot

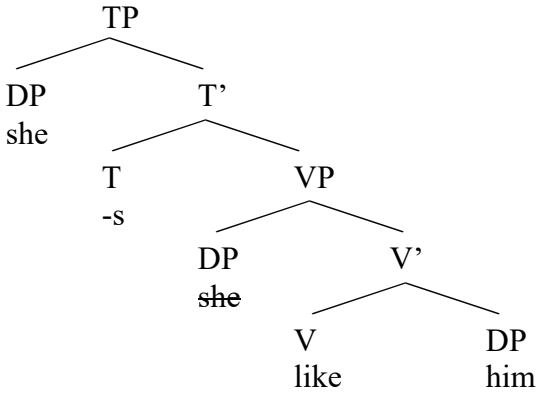
As the principles of GB were increasingly refined, the theory also became more complex. In lectures towards the end of the 1980's, Chomsky started asking whether the theoretical model could be rationalized in the sense of unifying, including eliminating, principles while maintaining, or possibly improving, empirical coverage (cf. Chomsky 2000, 93; Freidin and Vergnaud 2001, 642). This became the beginning of the Minimalist Program. Chomsky (1993) is the first paper that is written from an explicit minimalist perspective and one of the cases discussed there provides a helpful illustration of the underlying logic. Example (18) deals with case, which is not very conspicuous in English, but which can be seen in pronouns.

(18) a. Sara/John likes **him/her**.
b. **She/he** likes him/her.
c. Sara proved [**him** right]

As (18a) and (18b) show, pronoun subjects occur in nominative case and pronoun objects occur in accusative case, which of course are relics of the full-fledged morphological case system in older varieties of English. (18c) shows that accusative case can also occur in other environments. This raises the interesting question of how case assignment works – what are the configurations in which case licensing occurs?[ix] In Government and Binding, the structural representations underlying the bolded items in (18) are all different, though attempts had been made to use one principle (government) to account for all of them (Uriageka 1988, 2012). Before we can consider the argument, we need to do a brief detour to introduce some core concepts of phrase structure (see section 6.2 below, and Lasnik & Lohndal 2013 for a much more detailed treatment, including the historical development of phrase structure within generative grammar).

Consider the simplified structure in (19), where we are not illustrating the inner structure of the nominal phrase (called DP; see immediately below).

(19)



This is an instance of the X-bar schema (Chomsky 1970, 1986, Stowell 1981), a structural blueprint for syntactic structure. The schema contains a unified structure for each phrase: The phrase has a head (i.e., the phrase is endocentric), it has a complement as the sister of the head, and then a specifier is the 'subject' of the combined unit of head and complement. (19) has three phrases: the verb phrase (VP), the tense/agreement phrase (TP), and the determiner phrase (DP). If we apply this to VP, that means that V is the head of the verb phrase, *him* is the complement of the verb, and *she* is the specifier of the verb. The subject then moves to the canonical subject position in English, which is called the specifier position of T, or SpecTP for short (strikethrough marks that the constituent is not pronounced).[x] The tense and agreement morphology starts out as the head of TP and then gets unified in the morphology with the verb stem *like*. Now we are ready to go back to the problem posed by the data in (18).

In (19), *she* is in the specifier position of T, which is the locus of nominative case. Rather differently, *him* is the complement of the verb, which is to say that accusative case is licensed in that position. As for (18c), this is an exceptional case marking construction, since *him* is licensed in the specifier position of an embedded constituent. These environments for licensing case all look very different. Instead, Chomsky (1993) proposed that case licensing occurs in one type of configuration called a specifier-head configuration. Just like the subject position seems to be an agreement configuration (as subject-verb agreement shows), Chomsky proposes that there is a similar agreement configuration for the object, except that this agreement is not visible in languages like English. This makes it possible to say that case licensing of the accusative is very similar in (18a) and (18c) despite surface differences suggesting the opposite conclusion. That is, at some point in the derivation the object and verb must stand in a specifier-head relation so that Case can be assigned. Different variants of the view have proposed different points in the derivation where this would be achieved.[xi]

As Lohndal and Uriagereka (2014, 510) point out, this example makes the logic of the Minimalist Program clear: (i) Assume that the basic theoretical and empirical postulates of GB Case theory are correct: lexical items get case, certain syntactic heads carry the ability to provide case to a lexical item, and case licensing occurs in certain positions. (ii) The domains in which case licensing occurs are very different. (iii) A new theory offers a novel take on what a case licensing configuration is: specifier-head configurations, which are independently needed to account for subject-verb agreement in languages like English. Put differently, you pick a domain which is fairly well understood, you question some of the core parts of the analysis, and then you seek to develop a new and more principled and economic analysis (true to the Galilean ideal that has shaped much of Chomsky's work; see Allott, Lohndal & Rey this volume for further discussion).

Another area that Chomsky (1993) targeted concerned larger architectural aspects, namely the components of the grammar. In GB, the grammatical architecture had the structure

depicted in (17). This architecture contains a certain amount of overlap between D-structure and LF on the one hand, and S-structure and PF on the other hand (see Hornstein, Nunes, and Grohmann 2005 for much more discussion). D- and S-structure are clearly more grammar-internal than PF and LF, given that each sentence needs to receive a semantic encoding and a sound (or sign, as in the case of sign languages) encoding. Chomsky (1993) sets out to investigate whether these grammar-internal levels in the derivation could be eliminated. He reasons as follows:

> Each derivation determines a linguistic expression, an SD [Structural Description], which contains a pair (π, λ) meeting the interface conditions. Ideally, that would be the end of the story: each linguistic expression is an optimal realization of interface conditions expressed in elementary terms (chain link, local X-bar-theoretic relations), a pair (π, λ) satisfying these conditions and generated in the most economical way. Any additional structure or assumptions require empirical justification (Chomsky 1993, 4).

And since this paper, D-structure and S-structure have no longer been part of the grammatical component. Rather, the interfaces, sound and meaning, are regarded as the essential components, as Chomsky (2000, 91) argues:

> On these assumptions we understand L[anguage] to be a device that generates expressions Exp = <Phon, Sem>, where Phon provides the "instructions" for sensorimotor systems and Sem the "instructions" for systems of thought – information about sound and meaning, respectively, where "sound" and "meaning" are understood in internalist terms, "externalizable" for language use by the performance systems.

The emphasis on interfaces raises many new questions: What is the division of labor between the core syntax and the interfaces? How are the interfaces actually structured (see Uriagereka 2008, Samuels 2012, Lohndal 2014, among several, on the latter)?

Many other examples of minimalist reasoning could be provided as well. An excellent textbook exposition can be found in Hornstein, Nunes, and Grohmann (2005), and we refer the reader to this for more in depth discussion and a rich array of examples.

The minimalist reasoning is in part an application of Occam's Razor, what Martin and Uriagereka (2000) have labeled 'methodological minimalism':

> What one might call a "weak minimalist thesis" is nothing new. The drive for simple and nonredundant theories of the world (or Occam's razor) is taken for granted in the core sciences. Even within the more specialized science of linguistics, this working methodology has brought undeniable success. From such a perspective, minimalism is just a new way to refer to what many people have been doing for a long time: seeking the best way to theorize about a particular domain of inquiry. We think of this thesis as *methodological minimalism* (Martin and Uriagereka 2000, 1)

However, minimalism can also be viewed in a different sense, what Martin and Uriagereka (2000) label 'ontological minimalism'. As Chomsky (2000, 92) points out, progress has made it 'possible to consider some new questions about [the] F[aculty of]L[anguage]. In particular, we may ask the question, how well is FL designed?'. The idea is that language design may be optimal, 'approaching a "perfect solution" to minimal design specifications' (Chomsky 2000, 93). The latter is known as the Strong Minimalist Thesis (SMT). Chomsky is the first to stress that it would be surprising if the conclusion holds true, yet it would carry significant

implications if true. As such, the reduction and unification that characterize minimalism are also motivated by the SMT. It is not just a question of reducing or dispensing with grammar-internal levels, but with that comes the gain that the architecture only encompasses the virtually conceptually necessary interfaces, those that a theory of I-language cannot do without. This is all motivated by the quest to understand why certain principles of the language faculty hold and not others (Chomsky 2000, 92, Chomsky 2004). This type of why-question is characteristic of the Minimalist Program, and Chomsky (2000) emphasizes that they are extraordinarily difficult to answer:

> Questions of this kind are not often studied and might not be appropriate at the current level of understanding, which is, after all, still quite thin in a young and rapidly changing approach to the study of a central component of the human brain, perhaps the most complex object in the world, and not well understood beyond its most elementary properties (Chomsky 2000, 93).

As a consequence of this quest, other properties of the mind/brain have become more prominent (cf. Boeckx and Uriagereka 2006, 542). Chomsky (2005) emphasizes that there are in principle three factors that in some combination determine the adult language system: 'Assuming that the faculty of language has the general properties of other biological systems, we should, therefore, be seeking three factors that enter into growth of language in the individual' (Chomsky 2005, 6). These three factors are provided in (20).

(20) a. Genetic endowment specific to language
b. Experience
c. Principles not specific to the faculty of language

The first part is what is often referred to as Universal Grammar: it is whatever the genetic endowment is for acquisition of human language to take place. Furthermore, it is obvious that experience plays a role, given that a child growing up in Beijing acquires Mandarin Chinese and a child growing up in Athens acquires Greek. The last factor encompasses principles that are not specific to Universal Grammar. Chomsky argues that principles of data analysis may be among such principles, and also:

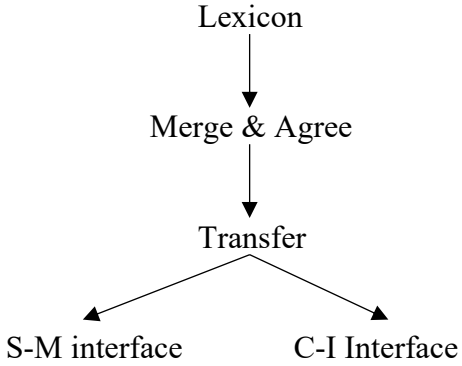
> […] principles of structural architecture and developmental constraints that enter into canalization, organic form, and action over a wide range, including principles of efficient computation, which would be expected to be of particular significance for computational systems such as language (Chomsky 2005, 6).

As Lohndal and Uriagereka (2017, 117) point out, these factors are very similar to the three factors that hold for organisms more generally (cf. Gould 2002, 259 on the 'adaptive triangle'). That said, it is not trivial to develop specific proposals for what the third factors actually are; see Lohndal and Uriagereka (2017) for comprehensive discussion. However, this focus on third factors has also led to a reduced emphasis on the role of Universal Grammar, hence, for example, Chomsky (2007) has the title 'approaching UG from below'. This is a marked shift compared to GB, where concerns about the size of Universal Grammar were non-existent. If a property is innate, it does not have to be somehow learned from the environment, so a big UG is compatible with taking arguments from the poverty of the stimulus seriously (see Crain, Giblin & Thornton this volume). In recent minimalism, the goal is to reach explanatory adequacy while attributing as little as possible to Universal Grammar. Another reason for this shift is the concern for biological and evolutionary adequacy which

characterizes minimalism. It is hard to see how all the separate parts (and their interrelationships) of the GB system could or would have evolved. How did the uniquely human ability for language arise in the species? The evolutionary literature is by now very rich and multiple proposals have been advanced (for reviews, see Fitch 2010, Reboul this volume). Even though minimalism does not offer a single answer to the question, the question itself has shaped a lot of research within minimalism, e.g., the increased emphasis on third factor effects.

The architecture of the grammar in Minimalism is illustrated in (21).

(21)



Let us consider this computational system in a bit more detail.

There is a lexicon, just as in GB and previous approaches. Based on whatever is in the lexicon, there is one basic mechanism that creates structure: Merge puts two items together, say, α and β, and that generates the set {α, β}, which typically then has α or β as its label: {α, {α, β} or {β, {α, β}}(Chomsky 1995). This, then, is the way to implement labelling in terms of (unordered) sets. More traditional tree structure representations are provided in (22).

(22)

Further items can be added and merged with α or β to generate binary-branching structures. This accounts for the creative linguistic ability that all humans possess. Merge comes in two different 'flavors': External Merge, which is the first time an item is merged into a structure, and Internal Merge, which means that, say, α is merged again later in the structure (hence accounting for displacement phenomena) (Chomsky 2004; see Svenonius this volume for extensive discussion). Importantly both ways of applying Merge are manifestations of the same operation. The Extension Condition is a principle which holds that all trees are extended at the top, meaning that new lexical items from the lexicon have to be merged on top of the existing structure, they cannot be 'tucked into' existing structure (Chomsky 1995).

In addition to Merge, there is at least the operation Agree, which is an abstract mechanism of dependency-formation. More specifically, the role of Agree is to ensure that syntactic features are checked or licensed. This is important in order to achieve descriptive adequacy, notably when it comes to constraining the grammar, since we know that dependency-formation is heavily constrained across languages. A simple example of dependency-formation when it comes to features can be seen in English subject-verb agreement as in (23).

(23) a. They are happy.
b. She is happy.

c. I am happy.

Subject-verb agreement is the morphological expression of an abstract syntactic relationship between the verb and the subject, since, as (23) illustrates, they have to agree in the features [NUMBER] (singular and plural) and [PERSON] (1, 2, 3). Agree is the mechanism that ensures that, say, the features of *they* match the features on the verb (see Adger 2003 for a textbook illustration). Notably, it is assumed that pronouns (and other nominals) carry number and person features that are interpreted at the C-I interface, whereas similar features on verbs are not interpreted by the C-I interface. For that reason, such features are uninterpretable, even though they can be morphologically realized, as in the case of English subject-verb agreement.

As structures are generated, they are at various points transferred to the two interfaces: the sensorimotor interface ('sound') and conceptual-intentional interface ('meaning'). That is, pieces of syntactic structure (called 'phases' within minimalism) are transferred to the interfaces and there given the necessary interpretation at both interfaces. Space does not allow us to elaborate on phases, but see Gallego (2012) for an extensive presentation.

The Minimalist Program is a program, meaning that it comes in different guises. As Chomsky (2000, 92) puts it: '[t]here are minimalist questions, but no minimalist answers, apart from those found in pursuing the program, perhaps that makes no sense, or that it makes sense but is premature'. Chomsky has never provided a more theoretical foundation of the notion of a program, but Boeckx (2006) does so by pointing to work by Lakatos (1968, 1970) and others. Lakatos argues that research programs have a core, a set of principles that the scholar has to adhere to. In addition, auxiliary assumptions are made, which often vary from one timeslice of the theory to another. The latter are often oversimplifications that the researcher knows are not true, but yet are made in order to be able to provide sufficiently detailed analyses of a set of phenomena. Because of this, Lakatos argues that the rigor of a research program may not at first be great, and that maturation is often slow. Relevantly for some of the criticism leveled against generative work, namely that there are so many counterexamples to proposed principles or generalizations, Lakatos has a specific view on the role of falsification: 'A corroborated falsifying hypothesis does not have sufficient power to enable a counterexample to eliminate a theory. If it had, we would eliminate all science instantly' (Lakatos 1968, 163). And he continues:

> A counterexample, in order to *reject*, to *eliminate* a theory, needs more powerful support than that which a lower-level falsifying hypothesis can provide: it needs the support of a theory with more corroborated content, with wide explanatory power. *There must be no elimination without the acceptance of a better theory* (Lakatos 1968, 163; his italics).

In this regard, Lakatos differs from Popper: '[…] in my conception criticism does not – and must not – kill as fast as Popper imagined. *Purely negative, destructive criticism, like 'refutation' or demonstration of an inconsistency does not eliminate a programme*' (Lakatos 1968, 183; his italics). This conception fits very well with how research within the Minimalist Program has taken place: It has been exploratory and continued despite obvious empirical or theoretical issues that needed to be fixed.

Boeckx (2006) and others argue that this programmatic nature of the Minimalist Program is a good thing. It allows researchers to pursue a multitude of approaches within the larger umbrella of minimalism. Put differently, it is possible to assume Lexicalism (e.g., Chomsky 1995: chapter 4, Williams 2007), Distributed Morphology (Embick and Noyer 2007, Embick 2015), cartography (Rizzi 1997, Shlonsky 2010), or an exo-skeletal approach

(Borer 2005a, b, 2013, Lohndal to appear). Within each version, a substantial body of work has emerged and the versions are developing fairly independently of each other. At the same time, it makes it hard to pin down exact criteria that can be used to label a particular scholar a minimalist. It also differs to what extent a particular scholar considers his or her work minimalist in nature, in particular in the ontological sense identified by Martin and Uriagereka (2000). However, it should be pointed out that some argue that minimalism should start producing what they label 'theories', here summarized by Lohndal and Uriagereka (2014, 520):

> Some scholars have argued that it is time to produce theories that can be evaluated and falsified (see, e.g., Hornstein 2009). Soon after the [Minimalist Program] was initiated, Lasnik (1999: 6) stated that 'there is not yet anything close to a Minimalist theory of language'. Differences in the underlying philosophy of science will in part determine whether a scholar regards programs as sufficient or whether theories are required for genuine progress (Lohndal and Uriagereka 2014, 520).

This raises obvious and thorny questions about what the difference between a program and a theory may be, issues that we cannot go into here.

That concludes our fairly brief overview of the general nature of the Minimalist Program. In the next section, we continue our discussion of this program by turning to some current tendencies.

6. Current tendencies

In this section, we will try to outline some of the current research trends within the Minimalist Program. A range of issues are just as important today as they have always been: Poverty of the Stimulus (see Crain, Giblin, and Thornton this volume), structure dependent rules, Universal Grammar, constraints on dependencies (see Müller this volume). Nevertheless, there are certain areas to which the searchlight is currently oriented. Some of these are discussed in other chapters, notably language evolution (Reboul this volume) and an increased emphasis on comparative grammar (Sheehan this volume). For that reason, we won't discuss these here. Rather, we will focus on three other topics: Features, the nature of phrase-structure representations (viz. labels and labeling), and extensions of the formal framework to new areas.

6.1. Features and the hierarchy of features

It has been pointed out in recent literature that there is a tension between Minimalism and the cartographic approach put forth by Rizzi (1997) and Cinque (1999). While Minimalism argues for a minimal role of Universal Grammar, the cartographic model assumes a very rich system of features, organized in one rich functional hierarchy. An oft-cited example is the following, which is Cinque's (1999: 106) proposal for a universal sequence of functional heads that host adverbials in their specifiers. The lexical items here are intended as examples of each category, and there is an Italian example (*tutto*) because there is no English example of this category.

(24) [*frankly* $\text{Mood}_{\text{speech act}}$ [*luckily* $\text{Mood}_{\text{evaluative}}$ [*allegedly* $\text{Mood}_{\text{evidential}}$ [*probably* $\text{Mod}_{\text{epistemic}}$ [*once* T(Past) [*then* T(Future) [*perhaps* $\text{Mood}_{\text{irrealis}}$ [*necessarily* $\text{Mod}_{\text{necessity}}$ [*possibly* $\text{Mod}_{\text{possibility}}$ [*usually* $\text{Asp}_{\text{habitual}}$ [*again* $\text{Asp}_{\text{repetitive(I)}}$ [*often* $\text{Asp}_{\text{frequentative(I)}}$ [*intentionally* $\text{Mod}_{\text{volitional}}$ [*quickly* $\text{Asp}_{\text{celerative(I)}}$ [*already* T(Anterior) [*no longer* $\text{Asp}_{\text{terminative}}$ [*still* $\text{Asp}_{\text{continuative}}$ [*always* $\text{Asp}_{\text{perfect(?)}}$ [*just* $\text{Asp}_{\text{retrospective}}$ [*soon* $\text{Asp}_{\text{proximative}}$ [*briefly* $\text{Asp}_{\text{durative}}$ [*characteristically(?)* $\text{Asp}_{\text{generic/progressive}}$ [*almost*

$\text{Asp}_{\text{prospective}}$ [*completely* $\text{Asp}_{\text{SgCompletive(I)}}$ [*tutto* $\text{Asp}_{\text{PlCompletive}}$ [*well* Voice [*fast/early* $\text{Asp}_{\text{celerative(II)}}$ [*again* $\text{Asp}_{\text{repetetive(II)}}$ [*often* $\text{Asp}_{\text{frequentiative(II)}}$ [*completely* $\text{Asp}_{\text{SgCompletive(II)}}$]]]]]]]]]]]]]]]]]]]]]]]]]]]]]

In opposition to this view, typically minimalist syntacticians assume a rather poor functional architecture, e.g. only the projections illustrated in (25).

(25) $[_{\text{CP}}$ C $[_{\text{TP}}$ T $[_{v\text{P}}$ *v* $[_{\text{VP}}$ V]]]]

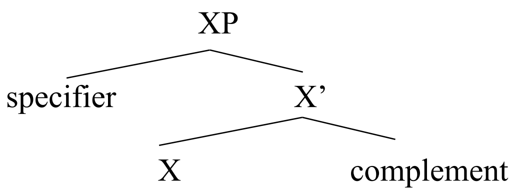
The CP domain is the discourse domain where e.g., topicalized constituents would move, as in *John, Mary really likes*. The TP domain typically hosts the (derived) subject and certain adverbs, wheras the vP and VP domains together are the domains for argument structure properties. However, frequently formulations such as the following can also be found: "C is shorthand for the region that Rizzi (1997) calls the "left periphery", possibly involving feature spread from fewer functional heads (maybe only one), […]" (Chomsky 2008, 9). This suggests that it is actually rather unclear what the size of the functional architecture across languages is, and how to capture the cross-linguistic generalizations that can be found.

Ramchand and Svenonius (2014) engage with this controversy (see also Wiltschko 2014). Their argument is that a fine-grained hierarchy of functional heads cannot be part of Universal Grammar as it does not fit with a minimal UG whose origin needs to be evolutionarily plausible. While such hierarchies emerge in some languages, Ramchand and Svenonius argue that these emerge in a highly constrained way. Specifically, there is a core tripartition of the clause into three domains, V, T and C (much like in (21)). Put differently, no language exists which has any different order, which is to say that V never appears above C in the hierarchy, and T never appears above C. In terms of why this hierarchy is the way it is, Ramchand and Svenonius argue that the tripartition has its source in extralinguistic cognition. As the authors state: "the most important source that we identify is grounded, we argue, in extralinguistic cognition: A cognitive proclivity to perceive experience in terms of events, situations, and propositions […]" (Ramchand and Svenonius 2014, 172). They continue by saying that "Granted, we have little direct evidence for these posited proclivities apart from the explananda themselves; but at present we do not know of plausible alternatives" (Ramchand and Svenonius 2014, 172). In years to come, it will be an important task to develop this approach further and better understand the relationship between the linguistic hierarchies and extralinguistic cognition.

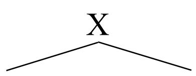
6.2. The nature of phrase structure representations: Labels and labeling

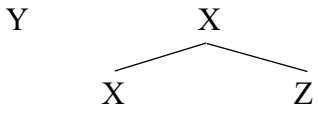
Across the history of generative grammar, the nature of phrase structure representations and how they are generated have always occupied a pivotal role (see e.g., Lasnik and Lohndal 2013). The development has gone from phrase structure rules to the X-bar schema in (26a) to the Bare Phrase Structure representation that emerged with the Minimalist Program (Chomsky 1995), shown in (26b).

(26) a.



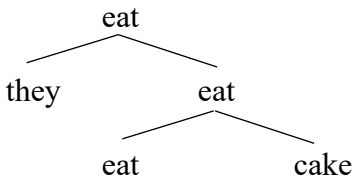
b.





Within X-bar theory, the positions were fixed: A specifier is a sister to X' (read: X-bar), and a complement is a sister to a head. In Bare Phrase Structure, all notions are relational, as can be seen in (26b) as all X-s are formally identical. The reasoning behind this shift was that phrase structure representations should not introduce anything beyond what is already in the lexicon, and bar-levels are not in the lexicon. Therefore, a verb phrase could look as in (27).[xii]

(27)

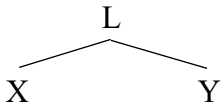


In (27), nothing beyond the lexical items are part of the representation and whether or not *they* is a specifier or a complement is something that is determined relationally (Chomsky 1995).

The change from X-bar theory to Bare Phrase Structure was seen as a major success, yet in the past 5-10 years, Chomsky has once again pushed the research frontiers by focusing on the labels themselves and on whether or not the algorithms need to incorporate facts about endocentricity or not. In previous models, each phrase had a head, thereby deriving endocentricity. In Chomsky's recent developments, endocentricity is still important for the syntactic derivation and to the interface systems, yet endocentricity is not a fundamental property of phrase structure qua phrase structure.

In deconstructing endocentricity, Chomsky has directed attention to the role of phrase structural labels: Are they needed (cf. Collins 2002), and if they are needed, at which level of the grammar may they be needed, i.e., what role do they play in syntactic derivations? As we have seen above, in Chomsky (1995), the operation Merge yielded a labeled syntactic structure. That is, Merge(X,Y) yields {L, {X, Y}} where L $\in$ {X, Y}.

(28)



However, adding the label is an extra assumption, Merge in its simplest form only takes two objects and puts them together: there is no labeling (Chomsky 2013, 2015, Chomsky, Gallego, and Ott 2019, Collins 2017). Thus it should be scrutinized carefully to see whether or not its existence is properly motivated. Chomsky (2013) argues that simplest Merge needs to be supplemented with another operation, which he calls Label. Essentially, this provides a way in which some of the effects of labelling can be accounted for, without postulating that constituents are ever labeled. This operation locates the 'structurally most prominent lexical item' within a syntactic object (Chomsky, Gallego, and Ott 2019, 247). For instance, if we have the syntactic object {H, XP}, where H is a lexical item and XP a complex object, then H will be the most prominent lexical item. The latter follows from H carrying a feature which can be located by the search algorithm as the most prominent item, which is to say that prominence is really determined by particular features. Since the search algorithm is always looking for the closest possible target, the features of H will be more prominent than the features embedded within the XP.

Chomsky and others have developed this approach to labeling in order to account for a variety of empirical patterns, ranging from movement, constraints on movement, and also diachronic change (see van Gelderen this volume). We cannot do justice to the rich technical details surrounding labels and labeling, except make the point that this has become a core area of current developments and that once again Chomsky has provided important contributions in shaping this research.

6.3. Extensions of the theory: Multilingualism

Chomsky (1965) defined the focus of most formal linguistics when he argued the following:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance (Chomsky 1965, 3).

Put differently, the monolingual speaker has been given the primary role in formal (and many non-formal) investigations of language (cf. Benmamoun, Montrul, and Polinsky 2013). Concretely, when a native speaker Y of language X is asked to provide judgments on strings of words, the researcher has deliberately ignored the fact that Y may also know other languages to various degrees. It has been argued that this kind of simplification made it possible to create new theories of hitertho unattested complex empirical patterns (Lohndal 2013).

There are several reasons why formal grammar would want to extend its empirical and theoretical scope. Speakers who know multiple languages to different degrees provide another type of data which in turn present new theoretical opportunities. They come in different guises, ranging from balanced multilingual speakers to receptive heritage speakers who are able to understand their heritage language but not produce it.[xiii] Given that a core objective in Chomskyan generative grammar is to model what a possible mental grammar is, data from multilingual speakers are extremely relevant (cf. Alexiadou and Lohndal 2016). These speakers also provide examples of what a possible grammar is, which is to say that the ecological validity of generative grammar is increased if it is able to account for what is after all the most common phenomenon today – namely that of being multilingual in one way or other (cf. Kupisch, Soares, Puig-Mayenco, and Rothman this volume).

With this as a backdrop, it should be added that Chomskyan generative grammar was extended fairly early on to situations and phenomena that do not fit the description in the quote above. Slabakova (this volume) shows this for second language learning, where substantial and crucial work especially started to emerge in the 1980's. However, multilingualism more generally has become much more prominent (cf. Kupisch, Soares, Puig-Mayenco, and Rothman this volume), in particular work on heritage languages (Polinsky 2018, Polinsky & Scontras 2020 make this vivid) and also work on code switching as an attempt to differentiate different formalisms. We will provide one example from the area of code switching and grammatical gender.

There has been a lot of discussion of how grammatical gender should be analyzed in the literature. Issues range from the nature of gender assignment to the syntactic location of gender features in the syntactic representation (see, among many others, Corbett 1991, Kramer 2016). The traditional assumption is that a gender feature is simply part of the lexical representation of a 'word'. However, recent theoretical work has questioned this idea and also the idea that words are the grammatically relevant unit. Instead, words are epiphenomenal, a result of the syntactic derivation. In particular, Distributed Morphology has promoted the idea

that the smallest units in a syntactic derivation are roots and functional elements that among others provide roots with a syntactic category (Marantz 1997, Alexiadou 2001, Arad 2003, Embick & Marantz 2008, Embick 2015). As an example, consider (29). In (29a), a root merges with a *v* to form a verb, and in (29b), the same root merges with an *n* to form a noun.

(29) a. $[_{v}\ v\ \sqrt{\text{ROOT}}\ ]$ b. $[_{n}\ n\ \sqrt{\text{ROOT}}\ ]$

A root has to be merged in order to ensure that a category is assigned. Embick and Marantz (2008: 6) label this *the categorization requirement*. In her comprehensive approach to grammatical gender, Kramer (2015) argues that *n* is the locus of gender assignment. If true, that means that gender is syntactically assigned. This has important ramifications for understanding language mixing involving grammatical gender. For instance, in American Norwegian, a heritage language of Norwegian spoken in the United States (see Haugen 1953), speakers easily assign grammatical gender to nouns borrowed from English, i.e., masculine, feminine, or neuter gender. (30) provides some examples, where the English items are in boldface (Grimstad, Riksem, Lohndal, and Åfarli 2018, 198).

(30) a. en **blanket**
a.M blanket
'a blanket'
b. ei **nurse**
a.F nurse
'a nurse'
c. et **crew**
a.N crew
'a crew'

As Grimstad, Riksem, Lohndal, and Åfarli (2018, 198) point out, the genders on English items are not identicial to the translational equivalents in Norwegian. The question is what determines gender assignment for English words in American Norwegian. Arguing that speakers have one lexical representation with, say, *blanket* carrying a masculine gender feature, and one without such a feature, simply pushes the problem back: How come speakers have acquired two lexical representations where one has grammatical gender and the other does not? Instead, Grimstad, Riksem, Lohndal, and Åfarli (2018) make the case that gender is assigned syntactically. Following Kramer's (2015) specific implementation, gender features would be on *n* (cf. (29b)), meaning that if *n* has a gender feature and the root is an English root, then we get e.g., *en blanket* as in (28a). The structures in (31) illustrate what this would look like for the three nouns in (30).

(31) a. $[_{n}\ n_{[\text{MASCULINE}]}\ \sqrt{\text{BLANKET}}\ ]$
b. $[_{n}\ n_{[\text{FEMININE}]}\ \sqrt{\text{NURSE}}\ ]$
c. $[_{n}\ n_{[\text{NEUTER}]}\ \sqrt{\text{CREW}}\ ]$

The major advantage of this analysis is that if gender were specified on the roots themselves, then multiple lexical entries for each root would be required. In contrast, the distributed approach in (31) allows the same root to be combined with any gender-flavor of *n*, and the data from American Norwegian support this variability since several speakers assign a different gender to the very same root. There are certain patterns in how gender assignment works, e.g., phonological similarity with Norwegian words has been argued by Haugen (1953) to be a factor. Space does not allow us to expand on this analysis any further, but it

demonstrates how data from multilingual speakers have become rather important in developing formal analyses of, say, grammatical gender and what the units of word formation actually are, see Alexiadou and Lohndal (2018) for further discussion.

7. Conclusion

The ongoing development of generative grammar is a testament to Chomsky's extraordinary ability to set the research agenda and continuously ask new questions and thereby sharpen our tools so that we get an ever better understanding of the unique human ability for language. In terms of the larger conceptual issues, there have been some changes over the years. An example is the change from Government and Binding to the Minimalist Program when it comes to approaching Universal Grammar: Do scholars assume that Universal Grammar is whatever it needs to be to account for the data, or do they instead assume that it should be as minimal as possible, in particular to get a possible handle on the evolutionary origin of the human ability for language? With the shift to the latter position, Chomsky has also positioned generative grammar better for interactions with research into cognition more generally.

References

Adger, D. 2003. *Core Syntax*. Oxford: Oxford University Press.

Alexiadou, A. 2001. *Functional Structure in Nominals*. Amsterdam: John Benjamins.

Alexiadou, A., and E. Anagnostopoulou. 1998. "Parametrizing AGR: word order, verb-movement and EPP-checking". *Natural Language and Linguistic Theory* 16: 491-539.

Alexiadou, A., and T. Lohndal. 2016. "Editorial: The Grammar of Multilingualism". *Frontiers in Psychology*, doi.org/10.3389/fpsyg.2016.01397

Alexiadou, A., and T. Lohndal. 2018. "Units of language mixing: A cross-linguistic perspective". *Frontiers in Psychology*, doi.org/10.3389/fpsyg.2018.01719

Arad, M. 2003. "Locality Constraints on the Interpretation of Roots: The Case of Hebrew Denominal Verbs". *Natural Language and Linguistic Theory* 21: 737-778.

Baker, M. 2008. "The microparameter in a microparametric world". In *The Limits of Syntactic Variation*, edited by T. Biberauer, 351-374. Amsterdam: John Benjamins.

Benmamoun, E., S. Montrul, and M. Polinsky. 2013. "Heritage Languages and Their Speakers: Opportunities and Challenges for Linguistics". *Theoretical Linguistics* 39: 129-181.

Boeckx, C. 2006. *Linguistic Minimalism*. Oxford: Oxford University Press.

Boeckx, C. 2013. *Syntactic Islands*. Cambridge: Cambridge University Press.

Boeckx, C., and J. Uriagereka. 2006. "Minimalism". In *The Oxford Handbook of Linguistic Interfaces*, edited by G. Ramchand and C. Reiss, 541-573. Oxford: Oxford University Press.

Borer, H. 2005. *Structuring Sense, Volume 1: In Name Only*. Oxford: Oxford University Press.

Borer, H. 2005b. *Structuring Sense, Volume 2: The Normal Course of Events*. Oxford: Oxford University Press.

Borer, H. 2013. *Structuring Sense, Volume 3: Taking Form*. Oxford: Oxford University Press.

Brame, M. K. 1976. *Conjectures and Refutations in Syntax and Semantics*. New York: Elsevier.

Chomsky, N. 1955. The Logical Structure of Linguistic Theory. Ms. Harvard University and MIT. [Revised version published in part by Plenum, New York, 1975]

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. 1970. "Remarks on nominalization". In *Readings in English*

Transformational Grammar, edited by R. A. Jacobs, and P. S. Rosenbaum, 184-221. Waltham, MA: Ginn.
Chomsky, N. 1973. "Conditions on transformations". In *A Festschrift for Morris Halle*, edited by S. Anderson, and P. Kiparsky, 232-286. New York: Holt, Rinehart and Winston.
Chomsky, N. 1975. *Reflections on Language*. New York: Pantheon.
Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.
Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.
Chomsky, N. 1993. "A Minimalist Program for Linguistic Theory". In *The View from Building 20*, edited by K. Hale, and S. J. Keyser, 1-52. Cambridge, MA: MIT Press.
Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
Chomsky, N. 2000. "Minimalist Inquiries: The Framework". In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, edited by R. Martin, D. Michaels, and J. Uriagereka, 89-155. Cambridge, MA: MIT Press.
Chomsky, N. 2001. "Derivation by Phase". In *Ken Hale: A Life in Language*, edited by M. Kenstowicz, 1-50. Cambridge, MA: MIT Press.
Chomsky, N. 2004. "Beyond Explanatory Adequacy". In *Structure and Beyond: The Cartography of Syntactic Structures*, edited by A. Belletti, 104-131. Oxford: Oxford University Press.
Chomsky, N. 2005. "Three Factors in Language Design". *Linguistic Inquiry* 36: 1-22.
Chomsky, N. 2007. "Approaching UG from Below". In *Interfaces + Recursion = Language?*, edited by H-M. Gärtner, and U. Sauerland, 1-29. Berlin: Mouton de Gruyter.
Chomsky, N. 2008. "On Phases". In *Foundational Issues in Linguistic Theory*, edited by C. Otero, R. Freidin, and M-L. Zubizarreta, 133-166. Cambridge, MA: MIT Press.
Chomsky, N. 2013. "Problems of Projection". *Lingua* 130: 33-49.
Chomsky, N. 2015. "Problems of Projection: Extensions". In *Structures, Strategies and Beyond*, edited by E. Di Domenico, C. Hamann, and S. Matteini, 1-16. Amsterdam: John Benjamins.
Chomsky, N. 2018. "Two Notions of Modularity". In *On Concepts, Modules, and Language: Cognitive Science at Its Core*, edited by R. G. de Almeida, and L. R. Gleitman, 25-40. Oxford: Oxford University Press.
Chomsky, N., Á. J. Gallego, and D. Ott. 2019. "Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges". *Catalan Journal of Linguistics* Special Issue: 229-261.
Chomsky, N., and H. Lasnik. 1977. "Filters and Control". *Linguistic Inquiry* 8: 425-504.
Chomsky, N., and H. Lasnik. 1993. "The theory of principles and parameters". In *Syntax: An International Handbook of Contemporary Research*, edited by J. Jacobs, A. von Stechow, W. Sternefeld, and T. Venneman, 506-569. New York: Walter de Gruyter.
Cinque, G. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press.
Collins, C. 2002. "Eliminating Labels". In *Derivation and Explanation in the Minimalist Program*, edited by S. D. Epstein, and T. D. Seely, 42-64. Malden: Blackwell.
Collins, C. 2017. "Merge(X,Y) = {X,Y}". In *Labels and Roots*, edited by L. S. Bauke and A. Blümel, 47-68. Berlin: Mouton de Gruyter.
Corbett, G. G. 1991. *Gender*. Cambridge: Cambridge University Press.

van Craenenbroeck, J., and J. Merchant. 2013. “Ellipsis phenomena”. In *The Cambridge Handbook of Generative Syntax*, edited by M. den Dikken, 701-745. Cambridge: Cambridge University Press.
D’Alessandro, R. 2015. “Null Subjects”. In *Contemporary Linguistic Parameters*, edited by A. Fábregas, J. Mateu, and M. T. Putnam, 201-226. London: Bloomsbury Press.
den Dikken, M., and A. Lahne. 2013. “The locality of syntactic dependencies”. In *The Cambridge Handbook of Generative Syntax*, edited by M. den Dikken, 655-698. Cambridge: Cambridge University Press.
Dougherty, R. C. 1973. “A survey of linguistic methods and arguments”. *Foundations of Language* 10: 432-490.
Embick, D. 2015. *The Morpheme: A Theoretical Introduction*. Berlin: Mouton de Gruyter.
Embick, D., and A. Marantz. 2008. “Architecture and Blocking”. *Linguistic Inquiry* 29: 1-53.
Embick, D., and R. Noyer. 2007. “Distributed Morphology and the syntax-morphology interface”. In *The Oxford Handbook of Linguistic Interfaces*, edited by G. Ramchand, and C. Reiss, 289-324. Oxford: Oxford University Press.
Fitch, W. T. 2010. *The Evolution of Language*. Cambridge: Cambridge University Press.
Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
Fodor, J. 2000. *The Mind doesn’t Work that Way*. Cambridge, MA: MIT Press.
Freidin, R. 2012. “A brief history of generative grammar”. In *The Routledge Handbook of the Philosophy of Language*, edited by D. Fara Graf and G. Russell, 895-916. London: Routledge.
Freidin, R., and H. Lasnik. 2011. “Some Roots of Minimalism”. In *The Oxford Handbook of Linguistic Minimalism*, edited by C. Boeckx, 1-26. Oxford: Oxford University Press.
Freidin, R., and J-R. Vergnaud. 2001. “Exquisite Connections: Some Remarks on the Evolution of Linguistic Theory”. *Lingua* 111: 128-172.
Gallego, Á. J. 2012. “Introduction: A framework of phases for linguistic theory”. In *Phases: Developing the Framework*, edited by Á. J. Gallego, 9-44. Berlin: Mouton de Gruyter.
Gould, S. J. 2002. *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.
Grimstad, M. Berg, B. Ramsevik Riksem, T. Lohndal, and T. A. Åfarli. 2018. “Lexicalist vs. exoskeletal approaches to language mixing”. *The Linguistic Review* 35: 187-218.
Haegeman, L. 1994. *Introduction to Government and Binding Theory*. Oxford: Blackwell.
Hankamer, J. 1973. “Unacceptable ambiguity”. *Linguistic Inquiry* 4: 17-68.
Harris, Z. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
Haspelmath, M. 2006. “Against Markedness (And What to Replace It With)”. *Journal of Linguistics* 42: 25-70.
Haspelmath, M. 2008. “Parametric versus functional explanations of syntactic universals”. In *The Limits of Syntactic Variation*, edited by Theresa Biberauer, 75-107. Amsterdam: John Benjamins.
Haugen, E. 1953. *The Norwegian Language in America*. Philadelphia: University of Philadelphia Press.
Hauser, M. D., N. Chomsky, and W. T. Fitch. 2002. “The Faculty of Language: What Is It, Who Has It, and How Did it Evolve?” *Science* 298: 1569-1579.
Hornstein, N. 2009. *A Theory of Syntax*. Cambridge: Cambridge University Press.
Hornstein, N. 2013. “Three Grades of Grammatical Involvement: Syntax from a

Minimalist Perspective". *Mind & Language* 28: 392-420.
Hornstein, N., J. Nunes, and K. K. Grohmann. 2005. *Understanding Minimalism*. Cambridge: Cambridge University Press.
Jaeggli, O., and N. Hyams. 1988. "Morphological uniformity and the setting of the null subject parameter". In *Proceedings of NELS 18*, edited by J. Blevins, and J. Carter, 238-253. Amherst: GLSA, University of Massachusetts, Amherst.
Judy, T., and J. Rothman. 2010. "From a superset ot a subset grammar and the semantic compensation hypothesis: subject pronoun and anaphora resolution evidence in L2 English". In *BUCLD 34: Proceedings of the 34th annual Boston University Conference on Language Development*, edited by K. Franich, K. M. Iserman, and L. L. Keil, 197-208. Somerville, MA: Cascadilla Press.
Kayne, R. S. 1980. "Extensions of Binding and Case-Marking". *Linguistic Inquiry* 11: 75-96.
Koizumi, M. 1995. "Phrase structure in minimalist syntax". Doctoral dissertation, MIT.
Kramer, R. 2015. *The Morphosyntax of Gender*. Oxford: Oxford University Press.
Kramer, R. 2016. "The location of gender features in the syntax". *Language and Linguistics Compass* 10: 661–677.
Lakatos, I. 1968. "Criticism and the Methodology of Scientific Research Programmes". *Proceedings of the Aristotelian Society* 69: 149-186.
Lakatos, I. 1970. "Falsifications and the methodology of scientific research programs". In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave, 91-195. Cambridge: Cambridge University Press.
Lakoff, G. 1970. "Global rules". *Language* 46: 627-639.
Lasnik, H. 1981. "Restricting the theory of transformations". In *Explanation in Linguistics*, edited by N. Hornstein, and D. Lightfoot, 152-173. London: Longmans.
Lasnik, H. 1999. *Minimalist Analysis*. Malden: Blackwell.
Lasnik, H. 2000. *Syntactic Structures Revisited: Contemporary Lectures on Classic Transformational Theory*. Cambridge, MA: MIT Press.
Lasnik, H. 2005. "Grammar, levels, and biology". In *The Cambridge Companion to Chomsky*, edited by J. McGilvray, 60-83. Cambridge: Cambridge University Press.
Lasnik, H. 2008. "On the Development of Case Theory: Triumphs and Challenges". In *Foundational Issues in Linguistic Theory*, edited by C. Otero, R. Freidin, and M-L. Zubizarreta, 17-41. Cambridge, MA: MIT Press.
Lasnik, H., and T. Lohndal. 2010. "Government-Binding/Principles & Parameters Theory". *Wiley Interdisciplinary Reviews: Cognitive Science* 1: 40-50.
Lasnik, H., and T. Lohndal. 2013. "Brief Overview of the History of Generative Grammar". In *The Cambridge Handbook of Generative Syntax*, edited by M. den Dikken, 26-60. Cambridge: Cambridge University Press.
Lasnik, H., and M. Saito. 1991. "On the subject of infinitives". In *Papers from the 27th Regional Meeting of the Chicago Linguistic Society*, edited by L. M. Dobrin, L. Nichols, and R. M. Rodriguez, 324-343. Chicago, IL: Chicago Linguistic Society.
Lasnik, H., and J. Uriagereka. 2005. *A Course in Minimalist Syntax*. Malden: Blackwell.
Lohndal, T. 2013. "Generative grammar and language mixing". *Theoretical Linguistics* 39: 215-224.
Lohndal, T. 2014. *Phrase Structure and Argument Structure: A Case Study of the Syntax-Semantics Interface*. Oxford: Oxford University Press.
Lohndal, T. To appear. "The exoskeletal model". In *The Cambridge Handbook of Distributed Morphology*, edited by A. Alexiadou, R. Kramer, A. Marantz, and I. Oltra-Massuet. Cambridge: Cambridge University Press.

Lohndal, T., and J. Uriagereka. 2014. “Principles and Parameters/Minimalism”. In *Routledge Handbook of Syntax*, edited by A. Carnie, Y. Sato, and D. Siddiqi, 509-525. London: Routledge.
Lohndal, T., and J. Uriagereka. 2017. “Third-factor explanations and Universal Grammar”. In *The Oxford Handbook of Universal Grammar*, edited by I. Roberts, 114-128. Oxford: Oxford University Press.
Marantz, A. 1997. “No escape from syntax: Don’t try morphological analysis in the privacy of your own lexicon”. *University of Pennsylvania Working Papers in Linguistics* 4: 201-225.
Martin, R., and J. Uriagereka. 2000. “Introduction: Some Possible Foundations of the Minimalist Program”. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, edited by R. Martin, D. Michaels, and J. Uriagereka, 1-29. Cambridge, MA: MIT Press.
Newmeyer, F. J. 2005. *Possible and Probable Languages*. Oxford: Oxford University Press.
Perlmutter, D. M. 1971. *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart & Winston.
Peters, S., and R. W. Ritchie. 1973. “On the generative power of transformational grammars”. *Information Sciences* 6: 49-83.
Polinsky, M. 2018. *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.
Polinsky, Maria, and Gregory Scontras. 2020. “Understanding Heritage Languages”. *Bilingualism: Language and Cognition* 23: 4-20.
Postal, P. M. 1974. *On Raising*. Cambridge, MA: MIT Press.
Ramchand, G., and P. Svenonius. 2014. “Deriving the functional hierarchy”. *Language Sciences* 46: 152-174.
Rizzi, L. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
Rizzi, L. 1997. “The fine structure of the left periphery”. In *Elements of Grammar*, edited by L. Haegeman, 281-337. Dordrecht: Kluwer.
Roberts, I. 2007. *Diachronic Syntax*. Oxford: Oxford University Press.
Ross, J. R. 1967. “Constraints on Variables in Syntax”. Doctoral dissertation, MIT. [Published as *Infinite Syntax!* Norwood, N.J.: Ablex, 1986.]
Ross, J. R. 1969. “Guess who?” In *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society*, edited by R. I. Binnick, A. Davison, G. M. Green, and J. L. Morgan, 252-639. Chicago Linguistics Society, University of Chicago.
Rothman, J. 2009a. “Pragmatic deficits with syntactic consequences? L2 pronominal subjects and the syntax-pragmatics interface”. *Journal of Pragmatics* 41: 951-973.
Rothman, J. 2009b. “Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages”. *International Journal of Bilingualism* 13: 155-163.
Rothman, J., and M. Iverson. 2007. “Input Type and Parameter Resetting: Is Naturalistic Input Necessary?” *International Review of Applied Linguistics* 45: 285-319.
Samuels, B. 2012. *Phonological Architecture: A Biolinguistic Perspectice*. Oxford: Oxford University Press.
Shlonsky, U. 2010. “The Cartographic Enterprise in Syntax”. *Language and Linguistics Compass* 4: 417-429.
Smith, N., and N. Allott. 2016. *Chomsky – Ideas and Ideals*. Third Edition. Cambridge: Cambridge University Press.
Stowell, T. 1981. “Origins of phrase structure”. Doctoral dissertation, MIT.
Taraldsen, K. T. 1980. “On the nominative island condition, vacuous application and the that-trace filter”. Bloomington, IA: Indiana University Linguistics Circle.

Uriagereka, J. 1998. “On Government”. Doctoral dissertation, University of Connecticut.
Uriagereka, J. 2008. *Syntactic Anchors*. Cambridge: Cambridge University Press.
Uriagereka, J. 2012. *Spell-Out and the Minimalist Program*. Oxford: Oxford University Press.
Vergnaud, J-R. 1977. “Letter to Noam Chomsky and Howard Lasnik on “Filters and Control”, April 17, 1977”. Printed in *Foundational Issues in Linguistic Theory*, 2008, edited by C. Otero, R. Freidin, and M-L. Zubizarreta, 4-15. Cambridge, MA: MIT Press.
Williams, E. 2007. “Dumping Lexicalism”. In In *The Oxford Handbook of Linguistic Interfaces*, edited by G. Ramchand, and C. Reiss, 353-382. Oxford: Oxford University Press.
Wiltschko, M. 2014. *The Universal Structure of Categories: Towards a Formal Typology*. Cambridge: Cambridge University Press.

Endnotes

[i] This chapter was written as part of the international research project *MultiGender* at the Centre for Advanced Study at the Norwegian Academy of Science and Letters in Oslo during the academic year 2019-2020, where both authors were fellows. We are grateful to an anonymous reviewer for very helpful comments on a previous version, and to Nicholas Allott for very detailed feedback. AL 554/8-1 (Alexiadou) is also hereby acknowledged.

[ii] Chomsky (1965, 4) emphasizes the importance of developing explicit theories. In doing so, he contrasts generative grammar with traditional grammars. The latter provide a lot of structural information, but they do not offer mechanisms that can account for the ‘regular and productive syntactic processes’. Such mechanisms are implicit as these traditional grammars presuppose that speakers and listeners have such knowledge. That is, speakers know where to put the finite verb in a *yes/no* question, but an explicit theory should explain the rule that governs how the finite verb *can* in *Can flying eagles swim?* becomes the first constituent. Importantly, trying to be explicit also raises new questions, such as why *Can eagles that fly swim?* is a question about the swimming ability of flying eagles, not the other way around. Such new questions have continuously led to novel generalizations about languages and language in general.

[iii] Since then, two additional questions have also been put on the agenda (cf. Boeckx 2006):

(i) a. How is that knowledge implemented in the brain?
b. How did that knowledge emerge in the species?

We will not discuss the first of these further here, but see Zaccarella and Trettenbein (this volume). We briefly discuss the latter in §5 below.

[iv] It is important to note that ‘knowledge’ incorporated into mental systems need not be true. However, when it comes to each person’s internal grammar, the question of truth does not really make sense. Furthermore, it is unclear whether knowledge needs to be believed. For sure, it does not have to be consciously available or integrated with general reasoning. Lastly, it does not have to be stored and retreivable (i.e. *represented* in the philosophers’ sense): It may be simply embodied/instantiated by the workings of a particular system (here the I-language). See Allott and Smith (this volume).

[v] Note that in principle the linguistic theory should help explain any linguistic behavior, not just intuitions. See Gross (this volume) for further discussion.

[vi] It should be pointed out that Hankamer (1973) argued that the constraint in question was universal, but given cross-linguistic variability in sluicing, that cannot be the case.

[vii] In the early stages of transformational grammar (Chomsky 1965), it was postulated that an *evaluation metric* helps the learner decide which grammar, amongst different ones which can all provide structural descriptions for the available primary linguistic data (PLD), is the optimal one. In terms of the evaluation metric, the core grammar is an optimal system.

[viii] This should not be taken to suggest that Fodor was committed to language being either a peripheral system or uninvestigable. Essentially, Fodor took for granted that the language faculty was essentially a database. See Allott & Smith (this volume) for much discussion.

[ix] In GB, a fundamental difference is introduced between case and Case, where the latter denotes what is called abstract Case; a syntactic Case feature which is not phonologically realized. An example would be *Sara* and *John* in (i), *Sara* is assumed to have abstract nominative Case whereas *John* has abstract accusative Case – despite there being no morphological marking of case.

(i) Sara loves John.

Abstract Case was first proposed by Vergnaud (1977) and further developed by Chomsky (1981).

Note also that the literature typically distinguishes between Case *assignment* and Case *checking*, a difference we set aside here, but see Lasnik (2008, 24) for an overview and a historical perspective.

[x] Evidence that this movement is necessary comes from facts about where adverbs and negation are placed. Semantically they modify the entire verb phrase, so they are inserted above the VP in the structure. Since the subject needs to precede the adverb, not follow it, as it would have if it did not move to SpecTP.

(i) a. She often likes him
b. She does not like him.

[xi] Implementations differ in terms of how this is done technically. Chomsky (1993) proposed covert movement at LF, whereas others have argued that there is displacement overtly of the object, but that the verb moves even higher (so as to precede the object on the surface) (Postal 1974, Lasnik and Saito 1991, Koizumi 1995, and much later work).

[xii] Setting aside higher functional material such as tense, and also any internal projection involving *they* and *cake*.

[xiii] Rothman (2009b, 156) defines a heritage language and a heritage speaker as follows:

> A language qualifies as a heritage language if it is a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society. […] [A]n individual qualifies as a heritage speaker if and only if he or she has some command of the heritage language acquired naturalistically […] although it is equally expected that such competence will differ from that of the native monolinguals of comparable age.