1 The Shift towards a Minimal UG

1.1 Introduction

The generative model focuses on the Faculty of Language as represented in the mind/brain. Early on in this framework (Chomsky 1957, 1965, 1975), the answer as to how children acquire language so effortlessly is seen as rooted in Universal Grammar (UG). Language acquisition, in this framework, is not imitation but an interaction between UG and exposure to a particular language. UG is "the system of principles, conditions, and rules that are elements or properties of all human languages not merely by accident but by necessity – of course, I mean biological, not logical, necessity" (Chomsky 1975: 29). However, although UG received a lot of attention, "principles of neural organization . . . deeply grounded in physical law" and the general "capacity to acquire knowledge" are acknowledged, even in the early period (Chomsky 1965: 59). Recently, these have been emphasized more, and have come to be known as third factor principles.

The premise of this book, and other work in diachronic generative syntax (e.g. van Gelderen 2019), is that the workings of third factor principles should be noticeable in language change as well. Since third factors bias the acquisition, a language learner may simplify the input in accordance with third factors resulting in more economic derivations. Languages also innovate in pragmatically motivated ways. Thus, language change involves a balancing act between economy and extravagance, the latter not constrained by third factors.

This introductory chapter will provide some background on the shift in emphasis towards third factors and some specifics on a few selected third factors, including those that receive the most attention in this book, determinacy and labeling. The remainder of the book looks at changes that can be argued to be prompted by third factors. The outline is as follows. In Chapter 2, I first outline the general shift that has taken place in generative grammar from language-specific principles to factors not restricted to the language faculty. In Chapters 3 and 4, I offer some background on two of the third factor principles, labeling and determinacy, respectively. Chapter 5 briefly discusses types of

change and Chapter 6 provides a justification of the data used. Chapter 7 presents a conclusion, an outline, and the major findings of the book.

1.2 From UG to Third Factor

This section sketches some of the thinking behind the Minimalist Program, putting less emphasis on UG and more on other factors, in particular third factors. The section also provides a discussion of a typical derivation and the interaction of the various principles, some of them third factor ones, those being the focus of this book.

1.2.1 Less UG

UG of the 1950s to 1970s includes a lot of language-specific instructions and rules. It initially includes rules for e.g. antecedent-reflexive relations, for passives, and for *wh*-movement. When the model shifts to the Principles and Parameters one, the parameters help to account for cross-linguistic differences while still keeping language acquisition central. In that model, children acquiring a language choose a setting (compatible with the data) for language-specific phenomena, such as *wh*-movement and null subjects. UG remains rich in the Government and Binding framework (Chomsky 1981: 3).

However, if the Faculty of Language developed in humans only 100,000 to 200,000 years ago – as is speculated – it makes sense to attribute less to language-specific principles. "UG must meet a criterion of evolvability" (Chomsky, Gallego, & Ott 2019: 230). If there were specific principles, they wouldn't have had enough time to evolve and, if there had been time, the changes would possibly have made languages very different from each other. That's why UG is currently assumed to contain just a simple operation merge that combines two elements into an unordered set, as in (1).

(1) Merge
Select two lexical items α and β and form the set {α, β} in a workspace. (Chomsky, Gallego, & Ott 2019)

Chomsky, Gallego, and Ott (2019) and Chomsky (2020: 34) refer to merge in the workspace as (capital) MERGE, to distinguish it from earlier instantiations although, as e.g. Krivochen (2020) has emphasized, there is no clear characterization of 'workspace'. For simplicity, I continue to use (noncapital) merge and use workspace in the sense of a phasal domain.

More important than principles specific to the language faculty, as in (1), are "general properties of organic systems" (Chomsky 2004: 105). In fact, three factors are considered crucial in the development of language, and they are listed below:

Three Factors

1. Genetic endowment, apparently nearly uniform for the species, which interprets part of the environment as linguistic experience, a nontrivial task that the infant carries out reflexively, and which determines the general course of the development of the language faculty. Among the genetic elements, some may impose computational limitations that disappear in a regular way through genetically timed maturation . . .
2. Experience, which leads to variation, within a fairly narrow range, as in the case of other subsystems of the human capacity and the organism generally.
3. Principles not specific to the faculty of language. (a) principles of data analysis that might be used in language acquisition and other domains; (b) 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)

The first factor principle, the "genetic endowment," comprises a greatly reduced UG which Chomsky (2005: 11–12) argues includes just merge, i.e. (1). Take two syntactic objects and combine them to make another object. Since merge can continue indefinitely, it also derives the recursive nature of language. Most linguists (e.g. Boeckx 2011) would add two other characteristics to UG, i.e. agree and phasal transfer (more about those later).

The second factor is the learner's need to be exposed to one or more languages (spoken or signed) to build the lexicon and to become familiar with interface constraints (again, more on these later). Lexical differences are responsible for all cross-linguistic variation, and parameters are now only relevant to that domain: speakers have to learn which features a lexical item has and which grammatical categories are included. There are some principles that help speakers acquire and use lexical and grammatical items and some of these can be formulated as in (2).

(2) Feature Economy
   (a) Utilize (innate) semantic features as the interpretable features connected to grammatical categories, if there is evidence for these. (van Gelderen 2011: 17)
   (b) If a specific feature appears more than once, one of these is interpretable and the others are uninterpretable. (Muysken 2008; van Gelderen 2013: 33)
   (c) Select feature bundles from the lexicon as economically as possible.

These are reminiscent of Longobardi's (2005) feature parameters, which also help the learner bundle the category and the features. However, the principles in (2) are formulated in terms of economy. Principle (2a) enables the acquisition of grammatical categories: a child needs to have lexical input for grammatical categories to appear. Chomsky (1995: 230; 381) suggests that "formal

features have semantic correlates and reflect semantic properties." Because a language learner uses innate concepts to assign semantic person and number features to, for instance, a noun, the learner will be able to hypothesize uninterpretable phi-features on another F (and will be able to bundle them there). Initially, a child uses lexical categories (as well as demonstrative pronouns) with interpretable features (as in Radford 2000) which are later used as uninterpretable features. The innate concepts (see also Pinker 1984: 244–245) involve, for instance, time, negation, and mass and are used linguistically as tense, negation, and number, respectively. Languages select certain categories to bear these, e.g. C or T for tense and D for definiteness. Principle (2b) enables the understanding of which features agree with each other and is relevant to C/T and v*/V, the phase heads, as well as possible other functional heads (as argued in Chapter 4). Principle (2c) makes it possible for features to be bundled in economical ways.

For syntactic change, and in particular grammaticalization, all the principles in (2) are relevant and work in tandem. (2a) is significant because a child will first acquire the semantic features of a lexical item but will, when there (no longer) is evidence for interpretable features, use the semantic features as interpretable. When lexical elements are renewed pragmatically, (2b) becomes relevant because there can only be one interpretable feature. As for (2c), a verb like *go* is listed with [motion] and [future] in the lexicon, and a selection would be possible for either the main or the auxiliary verbal use. If there is evidence in the data that *go* is used with another verb, the choice of features could just be [future]. The book is mainly about third factors and Feature Economy doesn't count as one. I do discuss it where relevant, e.g. in Chapter 4, Section 4.3.

The principles in (2) reference linguistic concepts and are therefore not third factors, although third factor principles, such as Biberauer's (2017) Maximize Minimal Means, shape them. Principle (3) is inspired by Abler (1989) and ultimately von Humboldt's (1836) insights.

(3) Maximize Minimal Means
   a. Make infinite use of finite means.
   b. Create what is not present in any of the associated constituents. (Humboldt 1836: 70, 67, from Biberauer 2017)

Biberauer outlines another set of clues that fit the second factor: for learners to notice silence, collocations, and recursion.

Third factor principles are not specific to language and include Minimal Search, Determinacy, and Structural Economy and take over some of what used to be attributed to first factor, i.e. UG, effects. In the next subsections, I briefly sketch the derivational model and give some examples of third factors, and will then devote the next two sections to the two third factors that will feature most prominently in the book.

1.2.2 *The Derivational Model*

The Minimalist Program proposes syntactic models and derivations that are very minimal and the same for every language. Interfacing with the syntactic derivation are the Conceptual-Intentional (CI) and Sensory-Motor (SM) systems. The former is responsible for providing an interpretation (e.g. through labeling) and includes nonlinguistic knowledge, whereas the latter is responsible for externalizing the derivation (i.e. providing a spoken or signed or written representation).

The minimalist model for deriving a sentence in the late 1990s involves making a selection from the lexicon, as in (4), and this is called a numeration.

(4) {they, read, will, the, books}

The selection procedure in (4) is to minimize the computational burden of keeping an entire lexicon active but, recently, a selection, lexical array, or numeration is no longer assumed (Chomsky, Gallego, & Ott 2019: 236–237). This means the entire lexicon is the input and the third factor principle Restrict Resources (more on this below) will restrict the options enough for the numeration not to be needed.

Using (1), the items from (4) or from the open lexicon are merged together, as in (5), from bottom to top.

(5) a. {the, books}
b. {read, {the, books}}
c. {they, {read, {the, books}}}
d. {will, {they, {read, {the, books}}}}
e. {they, {will, {<they>, {read, {the, books}}}}}

In steps (5ab), the object and the verb are combined, i.e. the v*P is constructed (ignoring the VP for brevity). The other steps involve the subject of the sentence being merged with the v*P in (5c) before the auxiliary *will* is in (5d). Sometimes the merge is 'internal', with input from inside the derivation, e.g. *they* in (5e), and I sometimes call that process movement, i.e. the subject moving to a higher position. When merge uses its own input, the term Internal Merge (IM) is used; otherwise External Merge (EM), as in (5abc).

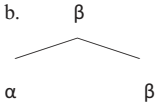
In (5e), I have represented the lower copy with angled brackets. This is to indicate that, at the level of externalization, it is not pronounced. In past work, linguists have also represented it with strike through, as in (6a), or with traces, as in (6b).

(6) a. {they, {will, {~~they~~, {read, {the, books}}}}}
b. They will t read the books.

In addition to indicating copies, angled brackets are also used for a special kind of merge, namely the pair-merge of adjuncts, as in (7a). As Chomsky (2004:

117–118) defines pair-merge, "there is also an asymmetric operation of adjunction, which takes two objects β and α and forms the ordered pair, α adjoined to β." In a tree, pair-merge looks like (7b).

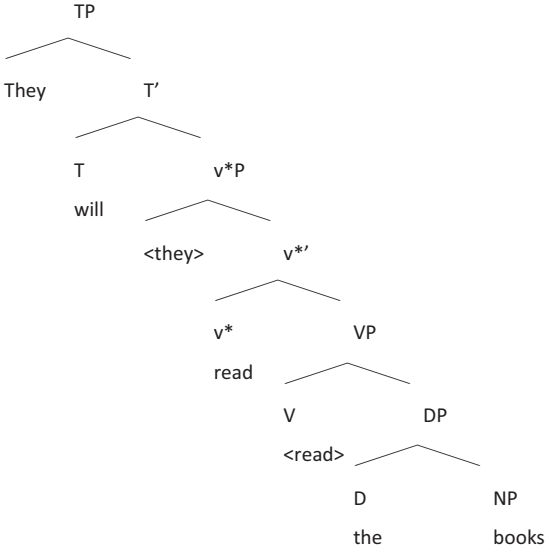
(7) a. Pair-Merge (α, β) = <α, β>

b.


Some authors use slightly different angled brackets, i.e. ⟨⟩ (Symbol font) for pair-merge, but I will not because the distinction between merge and copies will be clear without a special marker. Pair-merge ignores the adjoined element and is seen by many as problematic (e.g. Oseki 2015). In Chapter 6, I use pair-merge for VP adjuncts and discuss it a little more.

The curly brackets in (5) are used to indicate unordered sets, formed through set-merge, that need to be ordered when they are externalized at the SM Interface. As we saw, set-merge may be EM or IM and, as we'll still see, it is said to involve agree (cf. Kidwai & Mathew 2006).

Another way of representing the derivation in (5) is through a tree, as in (8). This structure uses an application of X′-structure that was current up to the mid-1990s. Although set notation, as in (5), is replacing tree structures in generative work, I follow others in using "trees, bracketing, and set notation interchangeably throughout" (Ott 2012: 9). The choice between trees and set notation will be based on clarity of exposition needs.

(8)


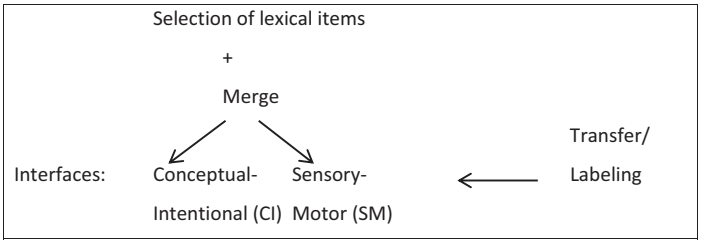


Figure 1.1 The Minimalist model of language generation

The v*P/VP layer is the theta domain, and the TP is the Tense Phrase, where information on finiteness and agreement is housed. In this tree, I have shown a split VP, with the Agent (the external argument) in the specifier of v*P and the Theme as the complement of the V. There is also a CP, not shown in (8), that indicates mood and tense and is situated above the TP.

The current Problems of Projection (PoP) approach within Minimalism (Chomsky 2013, 2015) insists that the use of the branches in (8) is deceiving because it suggests nodes where the branches come together that have labels. The PoP approach emphasizes that merge doesn't label. Syntax only combines objects and yields unordered sets {X, Y}, as in (5), without a label (Chomsky 2013: 42) and the labeling is done by a Labeling Algorithm when the syntax hands over its combined sets to the interfaces.

This brings us to the interfaces. After the selection of items from the lexicon and their merge into sets, the result is handed over (transferred) to the interfaces, as shown in Figure 1.1. The output from merge interfaces with the Sensory-Motor system and includes information on how to externalize the information (e.g. linearization). There are (at least) two possibilities to externalize the hierarchical structures. Kayne (1994) argues that what is higher in the tree (or least embedded in the merged structure) will also be spoken, signed, or written first. This is known as the Linear Correspondence Axiom. An alternative is to assume parameters in the Sensory-Motor system that linearize head first or head last, as elaborated on in Koeneman & Zeijlstra (2017: 224–227). The output from merge provides hierarchical structures to the Conceptual-Intentional system which the latter uses to compute meaning, events, and quantification. Many proposals exist for making the interface properties and the actual CI and SM systems more precise (e.g. Reinhart 2006 for the CI system).

1.2.3 Third Factors as Limits on Free Merge

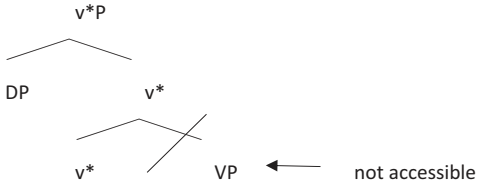
Having provided a brief sketch of a derivation, I now turn to some examples of third factors. Derivations are built from bottom to top, as in (5). With merge applying freely, they have to apply in stages, known as phases. A phase restricts the workspace so that the computational load is lessened. There are two phases in the sentential spine, namely CP and v*P. The v* is the head that introduces the external argument and there is a debate if structures without external argument, i.e. passives and unaccusatives, also constitute phases (see Chomsky 2001, 2008, who argues against that). I'll assume only v* counts, in addition to C.

Once the v* phase head is merged in (5c) and the v*P completed, the complement of v* is sent off to the interface (and the same is true once the CP is reached). Once a phase is transferred, it is no longer accessible to internal merge, according to a Third Factor Principle, the Phase Impenetrability Principle, as formulated in (9a) with an example in (9b).

(9) Phase Impenetrability Principle (PIC)

a. "The domain of H [v* or C] is not accessible to operations, but only the edge of HP [the head and specifiers]." (Chomsky 2004: 108)

b.



The reason for this formulation is to keep the notion of v*P and CP as phases but to allow their heads and specifiers to be accessible as escape hatches: the edge of the phase remains accessible at the next higher phase. Chomsky (2017, 2019) talks about restricting access to the input, Resource Restriction (see (13) below), and the PIC is one of these restricting devices. Recently, the PIC has been modified to only include internal merge and not agreement (Chomsky, Gallego, & Ott 2019: 241) but this remains a matter of discussion.

A practical example of the PIC for the *wh*-movement of an object is given in (10), where I ignore the TP for reasons of simplicity. In (10), the *wh*-object moves to the specifier of CP to express that the sentence is a question. Since v* is a phase head which transfers its complement to the interfaces, *what* must first move to the edge of the v*P phase, as in (10b), in order to be accessible for further movement to the CP. If it didn't move, it would be transferred together with the VP and be inaccessible under the PIC.

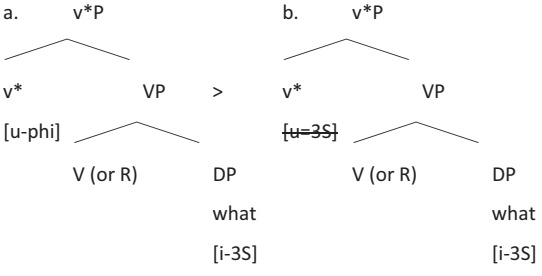
(10)

a. What did you eat?
b. CP[What did v*P[<what> v*P[you eat VP[<eat> <what>]]]]
v* = phase head IM of *what* to edge of phase
Transfer of VP upon completion of v*P

Thus, the PIC restates what locality principles do in earlier frameworks (e.g. subjacency), as we'll see, namely that movement must be local.

The lexical items in (5), (8), and (10), come with features. For instance, nouns have interpretable person and number features (phi-features) and v*, T, and C have uninterpretable ones. The uninterpretable features act as probes that need to find goals in their c-command domains to value the features (but there are many versions of how to achieve valuation). In (11), the phase head v* has uninterpretable phi-features, marked in the tree, which it checks with its object in-situ. This checking is done through Minimal Search in (12).

(11)



(12) Minimal Search
A head looks for something in a set/workspace, where 'something' can be a lower head, phrase, or features.

Wh-elements are also endowed with interpretable Q-features that are probed by a C with uninterpretable ones; I haven't shown these in (11), but we'll encounter some in Chapters 2–4.

Typically, only phase heads will have uninterpretable phi-features in structures like (11), so only v* and C. T inherits the phi-features from C but, for now, I continue to ignore T. Minimal Search is not "formally defined in the works of Chomsky" (Blümel 2017: 15) but is probably the most important of the third factors, at least in this book. Minimal Search is not only relevant to feature-checking in (11) but Labeling (see Section 1.3) and c-command (Chomsky 2005: 14) also use it and it is not restricted to just phase heads.

Three other Third Factor Principles are the Inclusiveness Condition (IC), the Extension Condition (EC), and the No Tampering Condition (NTC). All three are somewhat related and can be subsumed under the Principle of Resource Restriction, a formulation of which is given in (13).

(13) Principle of Resource Restriction
"[T]he operations should never extend the workspace." Chomsky (2019: 275)

The IC is given in (14) and it disallows traces and indices to indicate binding relations because these would be new objects not present in the original lexical array. The IC is also responsible for phrasal labels not being included through projection.

(14) Inclusiveness Condition (IC)
"[N]o new objects are added in the course of computation." (Chomsky 1995: 228)

The IC will be violated if expletives are allowed to merge late as a measure of last resort. In Chapter 5, I argue that that's not the case and that expletives are part of the original selection from the lexicon.

Not assuming a lexical numeration, as in (4), and leaving the lexicon open, as in Chomsky, Gallego, and Ott (2019), means the IC has to be reformulated and that is behind the Stability Principle in (15). Both (14) and (15) restrict the use of indices and traces, but (15) formulates that without a numeration.

(15) Stability Condition
"[T]he properties of a syntactic object can't change in the course of the derivation." (Chomsky 2019: 275)

Two related principles are the EC and NTC, as in (16) and (17). The EC states that trees are built from the bottom up and that merge applies to the top of the tree or to the output of previous merges. The NTC requires that the merge of "X and Y leaves the two SOs [Syntactic Objects] unchanged" (Chomsky 2008: 138), although that has to exclude internal merge. The NTC seems to include the EC.

(16) Extension Condition (EC)
"Merge always applies in the simplest possible form: at the root." (Chomsky 1995: 254)

(17) The No Tampering Condition (NTC)
Merge cannot make changes to the objects it affects.

Note that the EC is violated if heads adjoin to other heads in the syntax, as in (18), a point I briefly come back to in Section 1.3.1 and in more depth in Chapter 4 (Section 4.4).