The History of Syntax[1]
Peter W. Culicover

The history of thinking about and describing syntax goes back thousands of years. But from the perspective of theorizing about syntax, which is our concern here, a critical point of departure is Chomsky's *Syntactic Structures* (Chomsky 1957) henceforth *SS*.[2] I begin with some general observations about the goals of contemporary syntactic theory. Then, after briefly summarizing the main ideas of *SS*, and discussing methodology, I review some of the more important extensions, with an eye towards understanding where we are today, and how we got here.[3] I touch on some of the more prominent branch points later in the chapter, in order to preserve as much as possible a sense of the historical flow. For convenience, I refer to the direct line of development from *SS* as "mainstream" generative grammar (MGG). This term reflects the central role that the Chomskyan program has played in the field, both in terms of the development of his proposals and alternatives to them.

The contemporary history of syntax can be usefully understood in terms of a few fundamental questions. Answers to these questions have driven both the development of MGG, and the development of alternative syntactic theories. Among the questions that have proven to be most central and continue to fuel research are these:

- What is the nature of syntactic structure?
- What is the status within syntactic theory of grammatical functions, thematic roles, syntactic categories, branching structure, and invisible constituents?
- What is the right way to account for linear order?
- What is the right way to capture generalizations about relatedness of constructions?
- What is the explanatory role of processing in accounting for acceptability judgments and thus the empirical basis for syntactic theorizing?

1. GRAMMARS AND GRAMMATICALITY

A central assumption of MGG (and other theories) is that a language is a set of strings of words and morphemes that meet a set of well-formedness conditions. In MGG these are expressible as RULES. The rules constitute the grammar of the language, and are part of the native speaker's linguistic knowledge. One task of the linguist is to formulate and test hypotheses about what the rules of a language are, that is, to determine the grammar. The linguist's hypothesis and the native speaker's knowledge are both called the GRAMMAR.

The evidence for a child learning a language consists minimally of examples of expressions of the language produced in context. It is assumed that on the basis of this evidence the learner arrives at a grammar. The grammar provides the basis for the speaker to produce and understand utterances of the language.

The descriptive problem for the linguist is to correctly determine the form and content of the speaker's grammar. Since *Aspects* (Chomsky 1965) it has been assumed in MGG that the grammar is only imperfectly reflected in what a speaker actually says. Absent from the CORPUS of utterances is a vast (in fact infinite) amount of data that the speaker could produce, but hasn't produced, and could comprehend if exposed to it. It contains a substantial number of utterances that contain errors such as slips of the tongue, or are incomplete. Moreover, regular properties of the corpus such as the relative

frequency of various expressions and constructions may not be relevant to the grammar itself (in either sense), but to social and cognitive effects on the way in which the language defined by the grammar is used in communication.

The classical approach to discovery of the grammar has been to take the judgments of a native speaker about the acceptability of an expression to be a reflection of the native speaker's knowledge, that is, the grammar. In simple cases such an approach is very reliable. For instance, if we misorder the words of a sentence of a language such as English, the judgment of unacceptability is very strong, and reflects the knowledge of what the order should be. E.g., (1b) is ungrammatical because the article *the* follows rather than precedes the head of its phrase.

(1) a. The police arrested Sandy.
b. *Police the arrested Sandy.

Other cases are plausibly not a matter of grammar. Consider (2).

(2) a. Sandy divulged the answer, but I would never do it.
b. *Sandy knew the answer, but I would never do it.

Intuitively, the difference between the two sentences is that *do it* can refer only to an action, *divulge* denotes an action, while *know* does not. Since (2b) is ill-formed for semantic reasons, the burden of explanation can be borne by the semantics.[4]

The distinction between grammaticality and acceptability was highlighted by Miller and Chomsky 1963, who observed that a sentence can be well-formed in the sense that it follows the rules of linear ordering, phrase structure and morphological form, but is nevertheless unacceptable. Canonical cases involve center embedding (3).

(3) The patient that the doctor that the nurse called examined recovered.

The unacceptability of center embedding has been generally attributed to processing complexity, and not to grammar (Gibson 1998; Lewis 1997).

The distinction between grammaticality and acceptability has not played a significant role in syntactic theorizing until recently, primarily because of the unavailability of theories of the mechanisms (e.g. processing) other than syntax itself that could explain the judgments (see §7.4). The theoretical developments traced below are primarily anchored in the assumption that acceptability that cannot be attributed to semantics or pragmatics reflects properties of the grammar itself.

2. SYNTACTIC STRUCTURES AND THE STANDARD THEORY

2.1. Constituent structure

In *SS*, syntax is understood to be the theory of the structure of sentences in a language. This view has its direct antecedents in the theory of immediate constituents (IC), in which the function of syntax is to mediate between the observed form of a sentence and its meaning: "we could not understand the form of a language if we merely reduced all the complex forms to their ultimate constituents" (Bloomfield 1933:161). Bloomfield argued that in order to account for the meaning of a sentence, it is necessary to recognize how

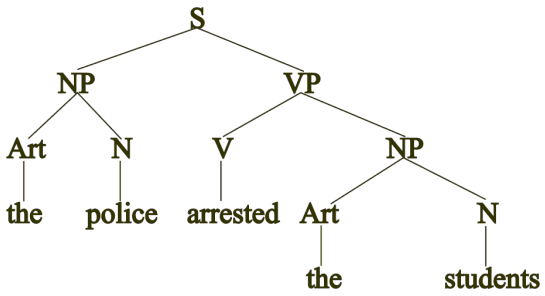
individual constituents (e.g. words and morphemes), constitute more complex forms, which themselves constitute more complex forms.

In *SS,* basic or KERNEL sentences were derived by the successive application of rewrite rules such as those in (4).

(4) S → NP VP
VP → V NP
NP → Art N
V → {arrested, ...}
Art → {the, a, ...}
N → {police, students, ...}

The application of such rules defines the IC structure of the sentence, e.g.,

(5)



2.2. Transformations

The fundamental innovation of *SS* was to combine IC analysis with Harris' observation (e.g. Harris 1951) that sentences with (more or less) the same words and meaning are systematically related. For example, the active and the passive, exemplified in (6), are essentially synonymous and differ only by the arrangement of the words and a few individual forms (*be,* the inflection on the main verb, *by*).

(6) a. The police arrested the students.
b. The students were arrested by the police.

For Harris, such relationships were captured through TRANSFORMATIONS of strings of words and morphemes.

In *SS*, such relationships among sentences are captured in terms of transformations of STRUCTURES. The passive transformation in *SS,* shown in (7), maps the structure of the active (e.g. (5)) into the structure of the passive. The object of the active, NP_2, occupies the subject position of the passive, and the subject of the active, NP_1, becomes the complement of the preposition *by*. A form of the verb *be* is inserted with the passive morpheme +*en*. A subsequent transformation attaches +*en* to the verb.

(7) (NP_1) V NP_2 ⇒ NP_2 be+en V (by NP_1)

Chomsky notes in *SS* that the passive construction has distinctive properties: the passive participle goes with *be*, a transitive passive verb lacks a direct object,[5] the agentive *by*-phrase may appear in the passive but not in the active, the exact semantic restrictions imposed on the object of the active are imposed on the subject of the passive, and the semantic restrictions on the subject of the active are imposed on the *by*-phrase. The passive could be described independently of the active, but such a description would be redundant and would not explicitly capture the relationship between the two constructions. Chomsky concludes (p. 43), "This inelegant duplication, as well as the special restrictions involving the element *be*+*en*, can be avoided ONLY [my emphasis – PWC] if we deliberately exclude passives from the grammar of phrase structure, and reintroduce them by a rule … ." Much of MGG and alternatives follow from responses to this conclusion.

Deriving the passive from the active by a RULE captures not only their synonymy, but also the distributional facts. Thus, Chomsky argued, phrase structure rules (PSRs) are not sufficient to characterize linguistic competence. A phrase structure characterization of the phenomena can capture the facts, but at the expense of generality and simplicity, as in the case of the English passive.

More complex sentences were derived in *SS* by the application of GENERALIZED TRANSFORMATIONS that applied to multiple simple sentences, as in (8).

(8) $\begin{bmatrix}\text{the police arrested the students}\\ \text{the students were protesting}\end{bmatrix} \Longrightarrow$

The police arrested the students who were protesting.

2.3. The shift to the Standard Theory

The shift from the *SS* theory to the Standard Theory (ST) in Chomsky 1965 was marked by three innovations: (i) since any order of application of the same rewrite rules produces the same structure, it is assumed in ST that PSRs such as (4a-c) specify a set of rooted trees as in (5) (Lasnik and Kupin 1977); (ii) since the full expressive power of generalized transformations is not needed, it was assumed in ST that complex structures are also specified by the PSRs, extended to allow for recursion, as in (9);

(9) S → NP VP
VP → V NP
NP → Art N
NP → Art N S

(iii) instead of rewrite rules, it was assumed that there is a LEXICON that specifies the properties of individual lexical items. A lexical item is inserted into a structure that is compatible with its properties, e.g. a transitive verb is inserted into a structure like (5) only if there is an NP in VP.

2.4. Levels of representation in the Standard Theory

Chomsky1965 proposed that there are two levels of syntactic representation of a sentence, DEEP STRUCTURE and SURFACE STRUCTURE, related by sets of transformations. The meaning of a sentence, in particular the assignment of THEMATIC (θ-)ROLES (e.g.

Agent, Patient) to the arguments, is determined by deep structure, while surface structure corresponds to the observed form, including linear order (now called PHONETIC FORM (PF)).

2.5. Constraining movement

A central consequence of the hypothesis that there are at least two transformationally related levels of syntactic representation is that constituents MOVE from their underlying positions to their observed positions in the structure. An example of movement is the derivation of the passive construction, in which the deep structure object moves to surface structure subject. Another is the movement of the English inflected auxiliary in subject Aux inversion (SAI) in (10b).

(10) a. Sandy will call.
b. Will Sandy ___ call.

Yet another example is seen in English wh-questions, where the interrogative phrase appears in a position distinct from the position that determines its syntactic and semantic function in the sentence (marked in (11) with underscore).

(11) What are you looking at __ ?

The question then arose, What kinds of movements are possible – how can they be constrained? Emonds 1970 observed that the passive transformation yields a structure that conforms to the general pattern of the language as characterized by the PSRs – that is, it is STRUCTURE PRESERVING. Emonds proposed that all transformations except those like SAI that apply to the highest level of the structure (the ROOT) are necessarily structure preserving. (In later developments, all transformations are assumed to be structure preserving.)

2.6. Long distance dependencies and island constraints

English wh-questions such as (11) exemplify a class of FILLER-GAP or A′ CONSTRUCTIONS in natural language. The wh-phrase is in an A′ position, that is, a position where its syntactic or semantic function is not determined. A′ positions contrast with A positions such as subject and direct object.

The contemporary analysis of A′ constructions in MGG posits a CHAIN that links the constituent in A′ position to a gap in the A position that defines its grammatical and semantic function. In what follows, the gap is marked with *t* co-subscripted with the constituent in A′ position. Thus (11) is represented as *What$_i$ are you looking at t_i*.

A distinctive characteristic of such constructions in languages like English is that there is no principled bound on the length of the chain. The wh-phrase may be linked to a gap in the complement, as in (12a), or in a more distant complement, as in (12b).

(12) a. Who$_i$ did you say [$_S$ you were looking at t_i]
b. Who$_i$ did you say [$_S$ everyone thinks … [$_S$ you were looking at t_i]]

The chain containing *who$_i$* and t_i is thus called a LONG DISTANCE DEPENDENCY (LDD).

The broad theoretical significance for syntactic theory of LLDs was recognized as early as Chomsky 1964. He observed that extraction of a wh-phrase from certain syntactic contexts is less than fully acceptable. Chomsky showed that while (13) is ambiguous, extraction of an NP corresponding to *the boy* as in (14) disambiguates – *walking to the railroad station* cannot be understood as a reduced relative modifying *the boy*. Chomsky concluded that extraction of *who* must be constrained in the structure (15).

(13) Mary saw the boy walking to the railroad station.
(14) Who did Mary see walking to the railroad station?
  a. 'Who did Mary see while she was walking to the railroad station?'
  b. Not: 'Who did Mary see who was walking to the railroad station?'
(15) Mary saw [$_{NP}$ [$_{NP}$ who] [$_{S}$ walking to the railroad station]

Chomsky's characterization of the configuration blocking extraction in (15) is that a phrase of category NP dominates another phrase of category NP, and the violation results from the extraction of the lower NP. He proposed "a hypothetical linguistic universal", subsequently referred to by Ross 1967:13 as the A-OVER-A PRINCIPLE (16).

(16) If [a] phrase X of category A is embedded within a larger phrase ZXW which is also of category A, then no rule applying to the category A applies to X (but only to ZXW).

Ross (1967) showed that the A-over-A principle does not account for the full range of restrictions on A′ extractions in English.[6] The configurations that inhibit extraction are called ISLANDS, and they are ruled out in MGG by ISLAND CONSTRAINTS. The reason why these must be expressed as constraints on rules (and not as rules of grammar themselves) is that the unacceptable examples are otherwise well-formed. For example, in a violation of the COMPLEX NP CONSTRAINT, as in (17b), the form of the relative clause is not problematic, since the relative pronoun is in the proper position. The problem is configuration of the chain.

(17) a. The police arrested the protesters who surrounded Sandy.
  b. *The person [$_{S}$ **who$_{i}$** the police arrested [$_{NP}$ the protesters [$_{S}$ who surrounded ***t$_{i}$***]]] was Sandy.

Moreover, the island constraints are arguably universal, and are thus not conditions on particular transformations.

Then the question arises how this knowledge could become part of a learner's grammar. Assuming that learners form grammars on the basis of the utterances they actually experience, it does not appear that there could be evidence that (17b) is ungrammatical, because it is well-formed from the perspective of structure (and rarely if ever produced). On the basis of such considerations, Chomsky (1965; 1973; 1981) argued that there are SYNTACTIC UNIVERSALS that constitute the human capacity for language. This is the ARGUMENT FROM THE POVERTY OF THE STIMULUS (APS), discussed further in §7.4.

3. UNIFORMITY

At this point it is helpful to consider a methodology of MGG that is responsible for much of its historical development. This methodology is UNIFORMITY (Culicover and Jackendoff 2005), which aims at eliminating redundancy in grammatical formulations.

3.1. Interface uniformity

INTERFACE UNIFORMITY (IU) is the assumption that sentences with the same meaning share a syntactic representation. If meaning is determined by deep structure, as in ST, sentences with the same meaning have the same deep structure representation. For example, the active and the passive are derived from the same representation, and the passive transformation does not affect their meaning. This point was generalized in MGG to the assumption that transformations in general do not add or change meaning (the Katz-Postal Hypothesis, Katz and Postal 1964).

Broad application of IU in the form of the Katz-Postal Hypothesis in the 1960s and early 1970s led to the emergence of Generative Semantics (GS). Consistent with ST, GS assumed two levels of syntactic representation, DS and SS. If transformations do not change meaning, all meaning must be determined at DS. Without a distinct syntactic level to represent logical form, GS concluded that DS was equivalent to the meaning. The decline of GS by the mid-1970s was propelled by a number of factors, most notably a failure to properly distinguish between genuinely syntactic and non-syntactic phenomena. Failure to distinguish in the theory among syntactic ill-formedness, semantic anomaly, presupposition failure, pragmatic infelicity, and so on, made it impossible to construct an explanatory account (see §3.1 and Katz and Bever 1976).

3.2. Structural uniformity

STRUCTURAL UNIFORMITY (SU) requires that if two constituents in two different sentences have the same grammatical function, then they have the same underlying representation. The canonical application of SU is to English wh-questions (18a) and similar A′ constructions, such as topicalization (18b) and relative clauses (18c).

(18) a. Who_i did the police arrest t_i?
b. Sandy_i, the police finally arrested t_i!
c. I was introduced to the person [$_S$ who_i the police arrested t_i].

Crucially, the filler has the same grammatical function as it (or a similar constituent) would have if it was in the position marked by the gap. SU thus REQUIRES that the filler occupy this position in deep structure. Classical MGG derivations apply MOVEMENT to map such a structure into one in which the filler is in the A′ position, forming a chain, as shown in (19).

(19) [the police finally arrested Sandy_i] $\Longrightarrow$ Sandy_i [the police finally arrested t_i]

3.3. Derivational uniformity

Assuming that at least some structures are derived by transformations, sameness of structure is captured by assuming DERIVATIONAL UNIFORMITY (DU). A typical case is sluicing, exemplified in (20).

(20) The police arrested someone, but I don't know who.

The second clause means 'but I don't know *who the police arrested*'. By IU, the two questions must have the same syntactic representation. By DU, since the full question has wh-movement (an MGG assumption), so must the sluiced case. So the derivation of (20) involves at least wh-movement of *who* and deletion of *the police arrested.* Similar reasoning is applied in the analysis of a variety of ELLIPTICAL constructions – see Culicover and Jackendoff 2005; 2012 for discussion.

4. THE EXTENDED STANDARD THEORY AND THE REVISED EXTENDED STANDARD THEORY

The Extended Standard Theory (EST) and the Revised Extended Theory (REST) are characterized by the recognition that some aspects of meaning are determined by derived structure (Jackendoff 1972), the introduction of the level of representation of LOGICAL FORM (LF), placing the burden of constraining the output of a transformation on general principles rather than the description of the transformation, constraining phrase structure through the X′ schema, and the introduction of TRACES.

4.1. S-structure and LF

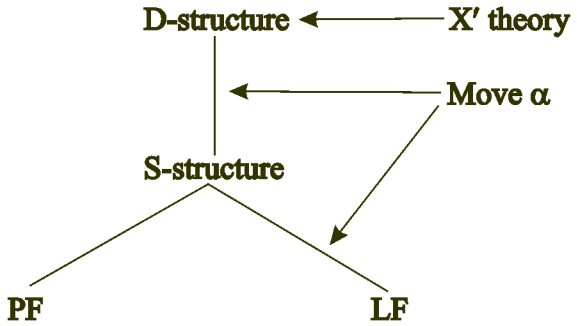
In REST, deep and surface structure are renamed D- and S-STRUCTURE. D-structure determines the thematic component of interpretation, through the assignment of θ-roles to arguments in their underlying canonical position. LF is the representation of the logical properties of a sentence that depend on its syntactic structure – it is determined by transformations that apply to S-structure (May 1985). For example, scope ambiguities such as (21) are derived by applying movement to the quantifier phrases to yield different hierarchical LF structures, as shown in (22). The traces *t* in (22) mark the canonical positions of the extracted arguments (§4.4).[7]

(21) Everyone speaks two languages.
  a. 'There are two languages that everyone speaks' [wide scope of *two languages*]
  b. 'Everyone is bilingual' [narrow scope of *two languages*]

(22) a. $[\text{two languages}]_j\ [_S\ \text{everyone}_i\ \text{speaks}\ t_j]$[8]
  b. $[\text{everyone}]_i\ [_S\ t_i\ \text{speaks}\ [\text{two languages}]_j]$

Characteristic of EST/REST is the T-MODEL of derivation in (23).

(23)



4.2. Move α

A transformation in ST was constrained by conditions stated in its formal description. Chomsky 1972 took a major step in shifting the burden of constraining transformations to general principles, by decomposing the passive into simpler movement operations that otherwise apply freely. Central to this step was the analysis of English nominalization. It was assumed since Lees 1960 that nominalizations such as (24a) should be derived from sentences (24b) on the basis of IU. The θ-roles are essentially the same: *the enemy* is the Agent, and *the city* is the Patient.

(24) a. the enemy's destruction of the city
b. The enemy destroyed the city.

Applying DU, Lakoff 1965 argued that the nominalization transformation should apply to the output of passive, on the basis of examples such as (25a,b).

(25) a. the city's destruction by the enemy
b. The city was destroyed by the enemy.

However, Chomsky showed that passive can be decomposed into a structure preserving movement of the object to the subject position in both S's and NPs, satisfying SU. (26) illustrates. ([$_{NP}$ *e*] denotes an empty NP position.)

(26) a. [$_{NP}$ *e*] destruction (of) <u>the city</u> (by the enemy) ⟹ [<u>the city</u>] *('s) destruction (*of) (by the enemy)
b. [$_{NP}$ *e*] (be) destroyed (*of) <u>the city</u> (by the enemy) ⟹ [<u>the city</u>] (be) destroyed (*of) (by the enemy)

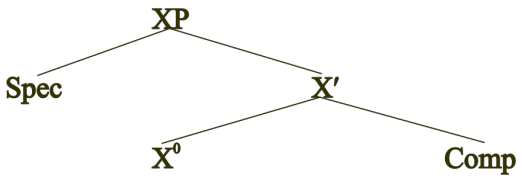
This analysis obviates the need for a nominalization transformation – the verb and its nominalization are lexically related. But more importantly, on this analysis the transformations do not need to be stated in terms of the properties of the syntactic structures to which they apply. Crucially, the only structural condition that the movement in (26), called Move α, must satisfy is that it is structure preserving, a general principle.

4.3. X′ theory

Virtually all syntactic theorizing has proceeded from the assumption that languages have words, that a word is a member of at least one LEXICAL CATEGORY, and that at least some phrases are projections of lexical categories (the HEADS) and acquire their categories from them.[9] Applying SU, MGG generalized the observed relationship between the structure of S and the structure of NP. The result was X′ theory (Chomsky 1972; Jackendoff 1977).

In the strongest form of X′ theory, every phrase of every category in every language has the structure in (27). X^0 is the HEAD of the phrase, Spec is the SPECIFIER, and Comp is the COMPLEMENT. Both Spec and Comp may be empty, or may consist of more than one constituent, depending on the selectional properties of the head.

(27)



X′ theory makes it possible to formulate a more uniform and constrained account of movement, on the assumption that all movement is structure preserving. Extending this view to wh-questions and inversion means that a landing site has to be found for the wh-phrase and for the inflected auxiliary. The wh-phrase must move to an available phrasal position, while the inflected auxiliary must move to an available head position. Chomsky 1981 proposed that the inflection (INFL=I^0) is the head of S (the projection is IP) and the complementizer C^0 is the head of CP. The structure is given by the PSRs in (28)-(29).

(28) IP → Spec I′
I′ → I^0 VP

(29) CP → Spec C′
C′ → C^0 IP

The wh-phrase moves to the specifier of C^0 and I^0 moves to C^0 in SAI; both movements are structure preserving.

(30) $[_{CP}\ [_{SPEC}\ e]\ C^0\ [_{IP}\ NP\ I^0_j\ \ldots\ \underline{\text{wh-XP}}_i\ \ldots]] \Longrightarrow$
$[_{CP}\ [_{SPEC}\ \underline{\text{wh-XP}}_i]\ I^0_j\text{+}C^0\ [_{IP}\ NP\ t_j\ \ldots\ t_i\ \ldots]]$

4.4. Traces

Wasow 1972; 1979 proposed traces as a way to solve a puzzling binding fact noted by Postal 1971.

(31) a. *He$_i$ thinks that Mary loves John$_i$.
b. *Who$_i$ does he$_i$ think that Mary loves t_i?

Wasow showed that (31b) can be ruled out by the same principle as (31a) if movement leaves a trace that is coindexed with the moved constituent; *John$_i$* and the trace of *who$_i$* bear the same syntactic relationship to *he$_i$*.

Traces were ultimately extended in REST and GB to all movements, so that the underlying structure is reflected in the derived structure. With gaps linked to extracted elements, it became possible to interpret sentences fully on the basis of surface structure. This fact has played a central role in the development of non-movement accounts of A′ constructions as early as Brame 1978 and Koster 1978, and was later given a comprehensive formal development in Head Driven Phrase Structure Grammar (HPSG) and Lexical Functional Grammar (LFG).

5. GOVERNMENT BINDING THEORY AND PRINCIPLES AND PARAMETERS THEORY

The shift to GB and PPT is characterized by the modularization of syntactic theory and by the introduction of the core-periphery distinction.

5.1. Modularity

MODULARITY is the idea that there are distinct components of the grammar, each of which obeys its own principles. Among the main components are: (i) X′ theory, θ theory, Case theory, binding theory, bounding theory, control theory and government theory. θ THEORY concerns the correspondence between syntactic structure and the θ-roles governed by a head. CASE THEORY regulates the movement of arguments to positions where they can be case-marked. BINDING THEORY concerns the syntactic relationships between referentially dependent elements (such as pronouns) and their antecedents. BOUNDING THEORY is a reformulation of the island constraints. CONTROL THEORY concerns the interpretation of verbal complements lacking overt subjects. GOVERNMENT THEORY regulates the functions of these component, e.g. θ-roles are assigned by a governor, Case is assigned by a governor, bounding constrains what a governor may govern, and the binding relation is constrained within a syntactic domain specified by a governor.

The core-periphery distinction holds that all languages share a common CORE GRAMMAR which is uniform up to parametric variation (e.g. the relative order of head and complement).

5.2. Extensions of X′ theory

Space precludes a review in detail of the main features of each of the components of GB theory, which are complex in their own right and in their interactions. Many of these grew out of earlier proposals and remain influential. To take just one case, Pollock 1989, in a very influential article, observed that English and French differ systematically in a number of respects, most notably that:

- the constituent that undergoes inversion in English questions must be a tensed auxiliary verb, while in French it may be a tensed main verb;

(32) a. English: He will go ⇒ Will he go?; He goes ⇒ Does he go?/*Goes he?

b. French: il va ⇒ va-t-il
he goes goes-t-he
"Does he go?'

- *not* in English follows an auxiliary verb, while in French negative *pas* follows a tensed main verb;

(33) a. English: He will not go. *He goes not.
b. French: Il (ne) va pas.
he NE goes NEG
"He doesn't go.'

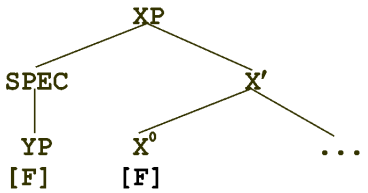
- adverbs in French immediately follow a tensed transitive main verb, while in English they follow the VP, not the verb.[10]

(34) a. English: John (often) kisses (*often) Mary.
b. French: Jean (*souvent) embrasse (souvent) Marie.
John often kisses often Mary

Pollock proposed that the main difference between English and French, then, is that in English, only auxiliary verbs attach to finite I^0, while in French main verbs do as well.

Analysis of additional details of verb-adverb ordering led Pollock to propose an "exploded" Infl, in which each feature is associated with a different head (AgrS, AgrO, and T(ense).[11] Extending Pollock's analysis (again following DU), Chomsky 1991; 1993 proposed that all movements to Spec positions are motivated by FEATURE CHECKING. A feature on a head is CHECKED or DISCHARGED if there is a constituent in its Spec that agrees with it, as in (35). If a feature is not checked, the resulting derivation is ill-formed.[12]

(35)

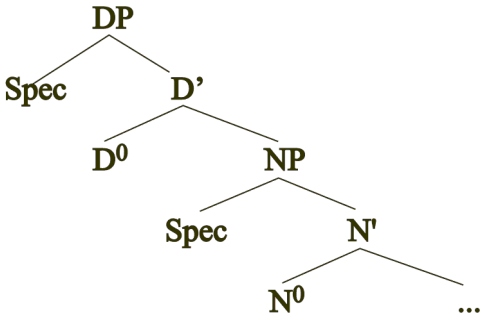


For example, in English a wh-phrase moves to Spec,CP in order to discharge the feature [WH] on C^0.

(36) $[_{CP}\ [_{SPEC}\ e]\ C^0[WH]\ [_{IP}\ NP\ I^0\ \ldots\ XP[WH]_i\ \ldots]] \Longrightarrow$
$[_{CP}\ [_{SPEC}\ XP[WH]_i]\ C^0[WH]\ [_{IP}\ NP\ I^0\ \ldots\ t_i\ \ldots]]$

Abney 1987 extended the application of functional categories to the NP and the parallelism with IP, arguing that N^0 is head of NP and the determiner is head of DP (37). DP is currently the standard notation in MGG for what was called NP in ST and EST.

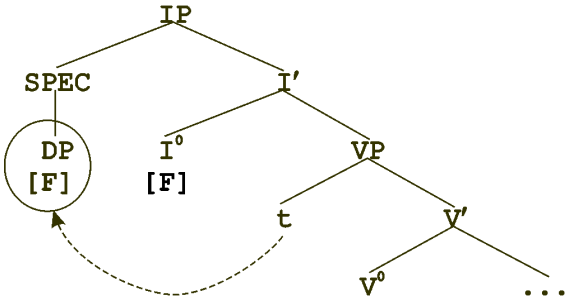
(37)



Abney also proposed that a possessive originates in Spec,NP and moves to Spec,DP to discharge a feature of D^0. This assumption allows all of the θ-roles of N^0 to be assigned within its maximal projection NP.

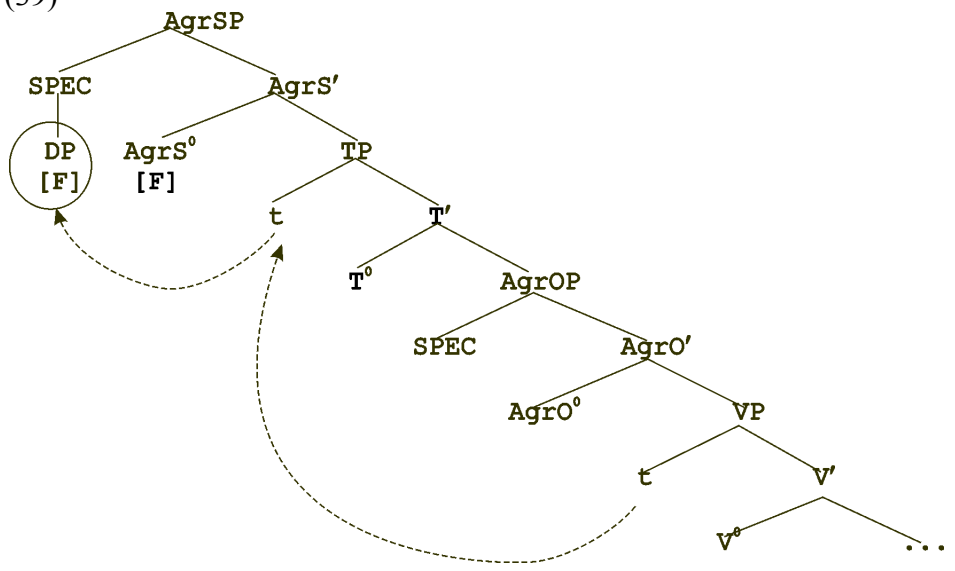
Extending this analysis to the sentence means that the subject DP originates as Spec,VP and moves to Spec,IP (see (38)).

(38)



This is the VP INTERNAL SUBJECT HYPOTHESIS. McCloskey 1997 argues that there is evidence for the internal subject positions predicted by the exploded Infl of Pollock, Chomsky and others, as shown in (39).[13]

(39)



Applying DU, the exploded Infl/feature checking analysis was extended in GB to the derivation of the passive. AgrS was assumed to be associated with a Case feature (as is AgrO in the transitive sentence). In the passive, the Case feature on the direct object is discharged if the direct object moves to Spec,AgrS.

5.3. Binding and movement

A major innovation in GB was to propose a tight interaction between binding and movement. Chomsky 1980 proposed that the distribution of referentially dependent elements such as pronouns and anaphors (such as reflexives and reciprocals) is governed by principles that have essentially the following content. Assume that coindexing of two expressions in the syntactic representation marks the coreference. Assume also that a constituent α BINDS a constituent β if α and β are coindexed and α C-COMMANDS β (i.e. if the constituent immediately dominating α dominates β). Then these principles hold: (A) A reflexive must be locally bound, (B) a pronoun cannot be locally bound, (C) a pronoun cannot bind its antecedent. Locality here is defined in terms of X′ theory – it is essentially the XP headed by a governor, with some extensions.

Constraining A movements was reconceived in GB in terms of conditions on the distribution of the trace in terms of binding theory. A movements are local, as is the binding of anaphors, suggesting that the trace of A movement is an anaphor. Non-local A movement violates Principle A of the binding theory since it is not locally bound.

5.4. Control

An important consequence of Uniformity is that it motivates abstract invisible constituents, which were introduced in REST and extended in GB/PPT. For example, the two sentences in (40) are synonymous.

(40) a. Susan_i expects [$_S$ that she_i will win].
b. Susan expects to win.

If *she* is coreferential with *Susan* (marked with coindexing), and *she* bears the role Winner, IU and SU lead to the conclusion that the structure of (40b) is (41), where PRO is an invisible pronominal (Chomsky 1981).

(41) Susan_i expects [$_S$ PRO_i to win].

Control theory is concerned with the distribution of PRO. Case theory plays a role in accounting for this distribution – PRO cannot be Case-marked. Government theory determines Case assignment – Case is assigned to a constituent that is governed. Hence PRO cannot be governed.

5.5. Antisymmetry

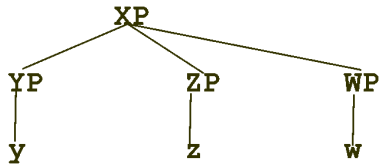
In MGG, from *SS* onward, linear order is represented explicitly in the PSRs. An influential proposal at the end of the GB/PPT era is antisymmetry theory (Kayne 1994). Kayne proposed to remove observed linear order as a grammatical primitive and to treat it as dependent on configuration. On Kayne's proposal, linear order is determined by the structure in (27) and the Linear Correspondence Axiom (42).[14]

(42) **Linear Correspondence Axiom (LCA)**
Let X, Y be non-terminals, and x, y terminals such that X dominates x and Y dominates y. Then if X asymmetrically c-commands Y, x precedes y. (Kayne 1994:33)

(43) **C-Command**
X c-commands Y iff X and Y are categories and X excludes Y and every category that dominates X dominates Y. (Kayne 1994:16)
Asymmetric C-Command
X asymmetrically c-commands Y iff X c-commands Y and Y does not c-command X. (Kayne 1994:4)

It follows from the LCA that underlying syntactic structure is uniformly binary branching and branches in the same direction (to the right, by stipulating "precedes" in (42)). Multiple branching precludes antisymmetry, and lack of antisymmetry results in no linear order by the LCA. Consider (44).

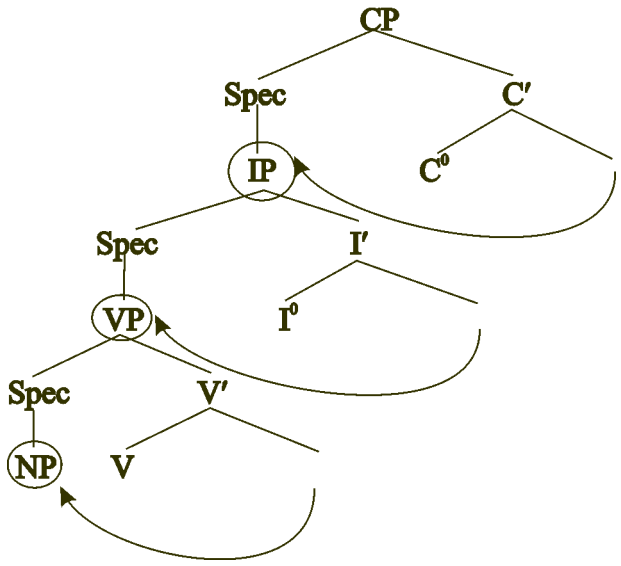
(44)



Since YP, ZP and WP c-command one another, there is no asymmetric c-command. Then there is no linear ordering defined between y, z and w. Hence XP is an impossible structure (according to the LCA).

Since observed linear order is often not of the form Spec-X^0-Comp, the LCA forces an account of many orders in terms of movement. For instance, in a verb-final language such as Japanese all complements and adjuncts of V must follow V and move to the left. On the additional assumption that all movement is structure preserving, there must be landing sites for all leftward movements. Moreover, there must functional heads whose features guarantee derivation of the overt order, by feature checking. A typical derivation is given in (45). The derived order NP-V-I^0-C^0 is that found in Japanese. Moving IP to Spec,CP correctly blocks wh-movement, which does not occur in Japanese and similar V-final languages.

(45)

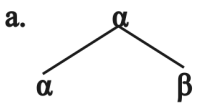


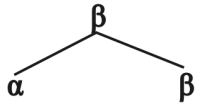
6. THE MINIMALIST PROGRAM

Chomsky's goal in the Minimalist Program (MP) is to reduce GB/PPT as much as possible to general principles of economy, to reduce derivations to their most primitive components, and to eliminate as much as possible the formal devices that had developed around the MGG approach (as summarized in the preceding sections). An accompanying goal is to investigate the extent to which natural language syntax can be viewed as deviating minimally from an ideal computational system.

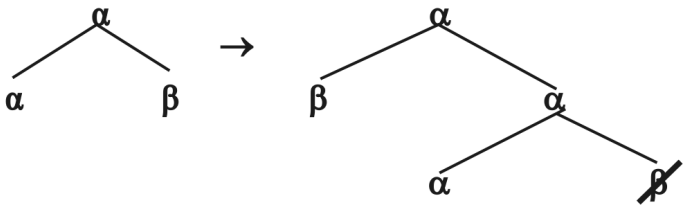
For instance, PSRs were eliminated in MP in favor of a primitive MERGE operation that combines two objects α and β, which are words or phrases, into a new

object; this operation is "external" Merge. The label of the resulting object is standardly assumed to be either that of α or β (46).

(46) a. 

b. 

Movement, or "internal Merge", is understood in MP as a combination of copying and deletion (47), an idea that has its roots in Chomsky's earliest work (Chomsky 1955).

(47) 

Movement is licensed when there is a feature of the merged constituent, the GOAL, that agrees with some feature of a head, the PROBE. The operation AGREE erases the feature of the goal if it matches that of the probe. Features that are not erased result in illegitimate derivations.

There are no island constraints in MP; rather, movement is constrained by DERIVATIONAL ECONOMY: each movement operation, or the length of each operation, or both, contributes to the complexity of a derivation (Chomsky 1995; Zwart 1996). Certain derivations can then be ruled out (in principle, at least) on the grounds that they are preempted by less complex derivations (see Johnson and Lappin 1997; 1999 for a critique). Suppose, for example, that α c-commands β, β commands γ, and both β and γ match the probe α. Since β is closer to α than γ, only β can Agree with α – this is the MINIMAL LINK CONDITION. For example, in (48) there are two wh-phrases but only the highest can undergo movement, as shown in (49).

(48) C^0[WH] [who[WH]$_i$ [$_{VP}$ was talking to whom[WH]$_j$]

(49) a. (I wonder) who[WH]$_i$ C^0[WH] [t_i [$_{VP}$ was talking to whom[WH]]

b. *(I wonder) whom[WH]$_j$ C^0[WH] [who[WH]$_i$ [$_{VP}$ was talking to t_j]]

This data falls under the SUPERIORITY CONDITION of Chomsky 1973.

Another consequence of the reformulation of derivations in terms of Merge in MP and the elimination of PSRs is that there are no levels of representation such as D-structure and S-structure. The phonological form of a phrase is specified as it is constructed by Merge, by the operation SPELL-OUT, and its interpretation is also assumed to be constructed dynamically in the course of the derivation.[15] An extension of this idea is found in DISTRIBUTED MORPHOLOGY, where a word is an abstract syntactic object, a complex word is derived by Merge, and the phonological form of a word is specified by Spell-out (Siddiqi 2010).

7. SOME CRITICAL BRANCH POINTS

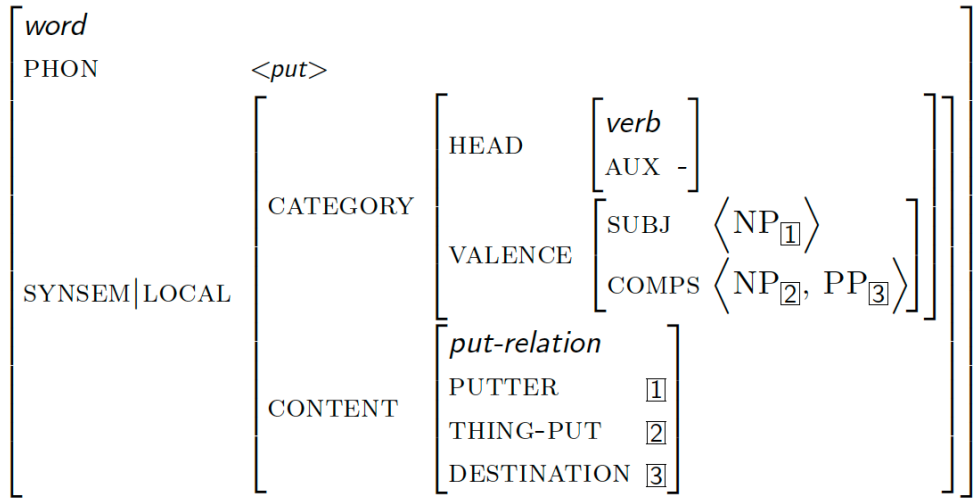
In this section I review briefly a number of theoretical developments that arose as a reaction to various aspects of MGG. A number of these have been taken up in the MP, resulting in some interesting convergences.

7.1. Derivations and representations

Head Driven Phrase Structure Grammar (HPSG; Pollard and Sag 1994) is one of the main alternatives to MGG. HPSG accounts for the derivation of sentences without assuming transformations. Since it assumes only one level of representation, HPSG is a MONOSTRATAL syntactic theory.

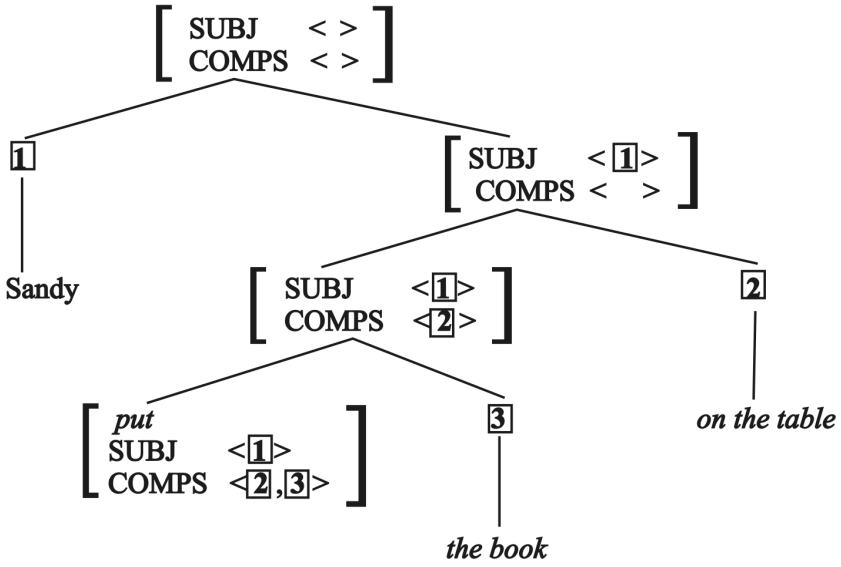
HPSG relaxes the assumption that phrase structure is a syntactic primitive whose properties are specified by a component of the grammar, leading to a major alternative to MGG.[16] In HPSG, phrase structure is implicit in the lexicon, hence it is a LEXICALIST theory. A typical lexical entry (for the verb *put*) is given in (50) (from Levine and Meurers 2006).

(50)



The hierarchical structure and linear ordering given by PSRs are expressed in HPSG by immediate dominance (ID) schemata and linear precedence (LP) statements. The features and the ID schema stipulate how to combine *put* with its complements (COMPS) and subject (SUBJ), and the semantic consequences. The structure emerges from this composition, as each VALENCE feature is satisfied by merging the current object with a phrase of a specified type according to the ID schemata. For example, *put* merges first with an NP, and the result is merged with a PP, to form the equivalent of a VP. (51) illustrates the derivation of the structure of *Sandy put the book on the table*.

(51)



HPSG deals with sentential organization differently from MGG (as do other approaches) , but there are important overlaps. In HPSG a phrase is a projection of its head, but there are no significant intermediate levels of syntactic representation, and no abstract functional heads – hence HPSG observes the VP internal subject hypothesis by default. There is feature checking, but the valence features in question correspond directly to observable syntactic properties of heads, e.g. SUBJ, COMPS, etc. Failure to SATURATE a valence feature means that it is passed up through the structure and must be satisfied at a later point in the derivation of the sentence, producing the possibility of chains. The VALENCE PRINCIPLE requires that the feature values of a phrase be identical to those of its head. For instance, if there is no direct object adjacent to the verb, then the valence feature is saturated by adjoining an NP to the left edge of the structure, which yields a filler-gap construction.

To the extent that comparable features are assumed in HPSG and MGG to license the arguments and their overt forms, the MGG approach can be seen to be a notational variant of the HPSG approach, differing primarily in that it has a more abstract structure, motivated through the application of DU.

7.2. Grammatical functions

From the beginning, MGG has taken the position that grammatical functions (GFs) such as subject and object are not primitives, but are defined in terms of "epistemologically prior" properties of utterances, e.g. linear precedence (Chomsky 1981:10). However, GFs are actually defined configurationally in MGG (Chomsky 1965:71): subject (of S) is an NP immediately dominated by S, object (of VP) is an NP immediately dominated by VP, predicate (of S) is a VP immediately dominated by S, and so on.

An important branch point in the development of syntactic theory is the assumption that the GFs are primitive. In LFG (Bresnan and Kaplan 1982) there is a level of F(UNCTIONAL)-STRUCTURE that corresponds in systematic ways to C(ONSTITUENT)-STRUCTURE. In a language such as English, the subject function corresponds to the configuration "NP immediately dominated by S", while in a case-marking language such

as Russian it corresponds to "NP marked with nominative case".[17] LFG's C-STRUCTURE is essentially equivalent to classical REST S-structure, and lacks functional heads except those that correspond to overt morphological forms.

In MGG, on the other hand, SU requires that the subject be represented uniformly, i.e. configurationally, across languages. Hence in an MGG analysis, the subject in a case-marking language such as Russian is in the same configuration as it is in English. Furthermore, by DU, if a particular configuration leads to a particular case-marking in one language, then it must lead to the same case-marking in all languages. Hence the subject in English has nominative case, etc. However, in English and many other languages there is clearly no morphological case. The solution in MGG is to assume that there is abstract Case (Chomsky 1980; 1981). Whether Case is realized morphologically is a secondary matter of spelling out.

Non-transformational theories refer explicitly to GFs to characterize the relationship between active and passive. The MGG account accomplishes this result by assigning Patient to the object, and then moving it to subject position. But in a non-transformational, or LEXICALIST account, Patient is assigned directly to the subject in virtue of the verb being in the passive form. In LFG, for example, f-structure plays a direct role in the analysis of the passive. There is a lexical rule that derives the passive verb from the active form. This rule reassigns the correspondences between the GFs and the θ roles governed by the verb (Bresnan 1982). The passive structure signals that the Agent role is not linked to the subject. Principles of mapping between syntactic structure and thematic representation then ensure that the Patient role is linked to the subject.

In HPSG a similar lexical rule rearranges the valence features of a verb and the θ-roles (Pollard and Sag 1987). Passive sentences are straightforward realizations of the basic structure of the language, similar to cases where the predicate is adjectival; cf. (52).

(52) a. Sandy was [$_{VP/AP}$ arrested by the police].
b. Sandy was [$_{AP}$ asleep at the wheel].

There is a lexical rule that remaps the roles to the syntactic arguments when the verb is passive. The rule applies generally to all verbs.

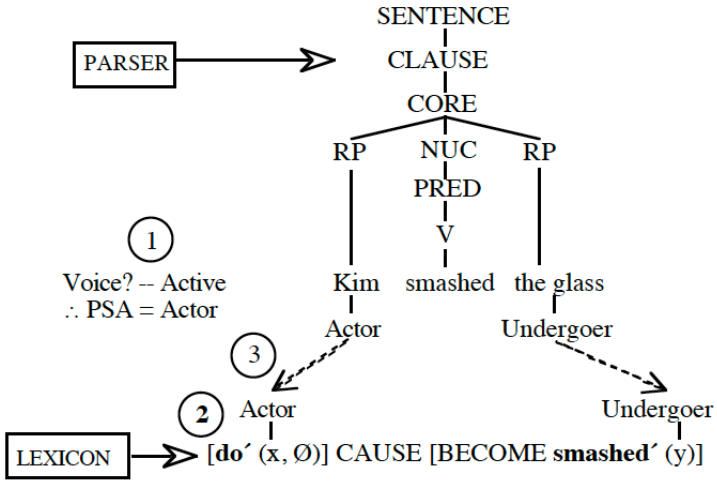
Similar devices are found in other approaches. The crucial mediating factor is the lexicon, where the θ-roles are associated with individual lexical items (Gruber 1972; Jackendoff 1972; 1983; 1990.

Relational Grammar (Blake 1990; Perlmutter 1983) takes the grammatical relations Subject, Direct Object, etc. to be syntactic primitives, rather than constituent structure. The structure is represented in terms of the assignment of grammatical relations to phrases, and constructions such as the passive are derived by reassigning the grammatical relations (e.g., underlying Subject is assigned to Direct Object, and underlying Object is assigned to an Oblique grammatical relation). Linear order is defined over the final stage (or STRATUM) of grammatical relation assignments.

In Role and Reference Grammar (Van Valin and LaPolla 1997) the syntactic representation is expressed not in terms of the classical syntactic categories (§4.3), but functional categories such as Clause, Referential Phrase, Predicate and so on.[18] This shift is motivated in part by research on less well-studied languages, where it is less clear that the generalizations can be captured in terms of the classical syntactic categories.[19] There

are no transformations in RRG; rather, there are rules that map directly between syntactic structure and semantic representations. Semantic arguments are ordered in a hierarchy according to their semantic role (Actor, Undergoer, etc.) and mapped to syntactic positions. An illustration is given in (53) (from van Valin Jr. 2010:736).

(53)



The PSA is the "privileged semantic argument", in this case the Actor, that gets mapped to the preverbal position – there is no mention of syntactic structure or grammatical functions such as Subject and Object.

In Simpler Syntax (Culicover and Jackendoff 2005) the Patient θ-role is linked to the Object GF. However, the passive is intransitive, so the Object cannot correspond to a syntactic constituent. It then corresponds by default to the Subject GF.

7.3. Constructions

MGG inherited the terminology CONSTRUCTION from traditional grammar; there is a "passive construction", a "wh-interrogative construction", and so on. By decomposing complex transformations such as the passive into the more primitive operation Move α, MGG gradually adopted the position that constructions as such are artifacts.

At the same time, many syntacticians have continued to treat constructions as grammatical primitives. Such a view has been explicitly formalized in Construction Grammar (Kay 2002), and has been widely adopted (see, e.g., Fillmore et al. 1988; Kay and Fillmore 1999; Goldberg 1995; 2006; Culicover and Jackendoff 2005; Sag to appear). The central empirical point is that some (if not all) syntactic structures have aspects of meaning associated with them that cannot be explained strictly in terms of the meanings of their constituents. In order to capture this part of the form-meaning relation, the construction *per se* must be part of the representation.

7.4. Explanation

While the Argument from the Poverty of the Stimulus (§1) is widely accepted, there are alternative views about where this type of knowledge could come from in the absence of direct experience. Processing accounts observe that there are well-formed examples with the same structure as the unacceptable examples, and attribute the judgment of ill-

formedness to processing complexity (typically characterized in terms of memory limitations) – see Hofmeister et al. to appear ; Hofmeister 2011; Sag et al. 2007 and for a critique Phillips to appear. Fodor 1978 and Hawkins 2004 have argued that grammars evolve to incorporate dispreferences for computationally complex structures. There are Bayesian approaches, which essentially argue that a structure can be judged unacceptable if there is an alternative structure that is significantly more likely, other things being equal (Pearl and Sprouse to appear).

8. CONCLUSION

The history of syntax shows a wide variety of theoretical alternatives, both internal to and external to MGG. Nevertheless, it is possible to see significant convergences, for example, the role of the lexicon in projecting syntactic structure and the dissociation of linear order and hierarchical structure. At the same time, there remain fundamental open questions, such as whether grammatical functions are primitives, the nature of syntactic categories, the relationship between syntactic structure and semantic interpretation, the status of syntactic constructions, the distinction between "core" and "peripheral" phenomena, the distinction between grammaticality and acceptability, and the origin and status of syntactic constraints. No doubt these issues will continue to animate the field, as they have since its inception.

RELATED TOPICS

Chapter 4
Chapter 5
Chapter 8
Chapter 9
Chapter 10
Chapter 11
Chapter 12
Chapter 13
Chapter 15
Chapter 21
Chapter 22
Chapter 24
Chapter 25
Chapter 26
Chapter 27
Chapter 30
Chapter 31
Chapter 32
Chapter 33

REFERENCES

Abney, S. (1987) *The Noun Phrase in its Sentential Aspect*. MIT.
Blake, B.J. (1990) *Relational grammar*. Routledge, London.

Bloomfield, L. (1933) *Language*. Holt, Rinehart & Winston, New York.
Brame, M. (1978) *Base Generated Syntax*. Noit Amrofer, Seattle, Washington.
Bresnan, J. (1982) The passive in grammatical theory. In *The Mental Representation of Grammatical Relations,* (Ed, Bresnan, J.) MIT Press, Cambridge, MA, pp. 3-86.
Bresnan, J. & Kaplan, R. (1982) Lexical functional grammar: a formal system for grammatical representations. In *The Mental Representation of Grammatical Relations,* (Ed, Bresnan, J.) MIT Press, Cambridge, MA., pp. 173-281.
Carnie, A. (2010) *Constituent structure*. Oxford University Press, Oxford.
Chomsky, N. (1955) *The Logical Structure of Linguistic Theory*. Plenum, New York.
Chomsky, N. (1957) *Syntactic Structures*. Mouton, The Hague.
Chomsky, N. (1964) *Current Issues in Linguistic Theory*. Mouton, The Hague.
Chomsky, N. (1965) *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
Chomsky, N. (1972) Remarks on nominalization. In *Readings in English Transformational Grammar,* (Eds, Jacobs, R. & Rosenbaum, P.) Ginn and Co., Waltham, Mass, pp. 184-221.
Chomsky, N. (1973) Conditions on Transformations. In *Festschrift for Morris Halle,* (Eds, Anderson, S. & Kiparsky, P.) Holt, Rinehart and Winston, New York, pp. 232-286.
Chomsky, N. (1980) On binding. *Linguistic Inquiry*, **11**, 1-46.
Chomsky, N. (1981) *Lectures on Government and Binding*. Foris Publications, Dordrecht, Holland.
Chomsky, N. (1991) Some notes on economy of derivation and representation. In *Principles and parameters in comparative grammar,* (Ed, Freidin, R.) MIT Press, Cambridge, MA, pp. 417-454.
Chomsky, N. (1993) A minimalist program for linguistic theory. In *The View from Building Twenty,* (Eds, Hale, K. & Keyser, S.J.) MIT Press, Cambridge, MA, pp. 1-52.
Chomsky, N. (1995) *The Minimalist Program*. MIT Press, Cambridge, MA.
Cinque, G. (1999) *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford University Press, Oxford.
Culicover, P.W. & Jackendoff, R. (2012) A domain-general cognitive relation and how language expresses it. *Language*, **88**, 305-340.
Culicover, P.W. & Jackendoff, R. (2005) *Simpler Syntax*. Oxford University Press, Oxford.
Emonds, J. (1970) *Root and Structure Preserving Transformations*. Indiana University Linguistics Club, Bloomington, Indiana.
Fillmore, C.J., Kay, P. & O'Connor, M.C. (1988) Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language*, **64**, 501-539.
Fodor, J.D. (1978) Parsing strategies and constraints on transformations. *Linguistic Inquiry*, **9**, 427-473.
Gibson, E. (1998) Linguistic Complexity: Locality of Syntactic Dependencies. *Cognition*, **68**, 1-76.
Goldberg, A.E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago.
Goldberg, A.E. (2006) *Constructions at Work: Constructionist Approaches in Context*. Oxford University Press, Oxford.

Graffi, G. (2001) *200 years of syntax: a critical survey*. John Benjamins Publishing Company, Amsterdam.
Gruber, J.S. (1972) *Functions of the Lexicon in Formal Descriptive Grammars*. Indiana University Linguistics Club, Bloomington, IN.
Harris, Z. (1951) *Methods in Structural Linguistics*. University of Chicago Press, Chicago.
Hawkins, J.A. (2004) *Complexity and Efficiency in Grammars*. Oxford University Press, Oxford.
Hofmeister, P., Staum Casasanto, L. & Sag, I.A. (to appear) Islands in the grammar? Standards of evidence. In *Experimental Syntax and Island Effects,* (Eds, Sprouse, J. & Hornstein, N.) Cambridge University Press, Cambridge.
Hofmeister, P. (2011) Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, **26**, 376-405.
Jackendoff, R. (1972) *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, MA.
Jackendoff, R. (1977) *X-Bar Syntax: A Study of Phrase Structure*. MIT Press, Cambridge, MA.
Jackendoff, R. (1983) *Semantics and Cognition*. MIT Press, Cambridge, MA.
Jackendoff, R. (1990) *Semantic Structures*. MIT Press, Cambridge, MA.
Johnson, D.E. & Lappin, S. (1997) A Critique of the Minimalist Program. *Linguistics and Philosophy*, **20**, 273-333.
Johnson, D.E. & Lappin, S. (1999) *Local Constraints vs. Economy*. CSLI, Stanford, CA.
Katz, J.J. & Bever, T.G. (1976) The rise and fall of empiricism. (Eds, Bever, T.G., Katz, J.J. & Langendoen, D.T.) Crowell, New York, pp. 11-64.
Katz, J.J. & Postal, P.M. (1964) *Toward an Integrated Theory of Linguistic Descriptions*. MIT Press, Cambridge, MA.
Kay, P. (2002) An Informal Sketch of a Formal Architecture for Construction Grammar. *Grammars*, **5**, 1-19.
Kay, P. & Fillmore, C.J. (1999) Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language*, **75**, 1-33.
Kayne, R.S. (1994) *The Antisymmetry of Syntax*. MIT Press, Cambridge, MA.
Koch, K. & Matthewson, L. (2009) The lexical category debate in Salish and its relevance for Tagalog. *Theoretical Linguistics*, **35**, 125-137.
Koster, J. (1978) *Locality Principles in Syntax*. Foris Publications, Dordrecht, Holland.
Lakoff, G. (1965) *Irregularity in Syntax*. Holt, Rinehart and Winston, New York.
Lasnik, H. & Kupin, J.J. (1977) A restrictive theory of transformational grammar. *Theoretical Linguistics*, **4**, 173-196.
Lees, R.B. (1960) The grammar of English nominalizations. Indiana University Research Center in Anthropology, Bloomington, Indiana.
Levine, R.D. & Meurers, W.D. (2006) Head-Driven Phrase Structure Grammar Linguistic Approach, Formal Foundations, and Computational. In *Encyclopedia of Language and Linguistics, 2nd Edition,* (Ed, Brown, K.) Elsevier, Oxford.
Lewis, R.L. (1997) Specifying architectures for language processing: Process, control, and memory in parsing and interpretation. In *Architectures and Mechanisms for Language Processing,* (Eds, Pickering, M. & Clifton, C.) Cambridge University Press., Cambridge.

Matthews, P.H. (1993) *Grammatical theory in the United States from Bloomfield to Chomsky*. Cambridge University Press,
May, R. (1985) *Logical Form*. MIT Press, Cambridge, MA.
McCloskey, J. (1997) Subjecthood and subject positions. In *Elements of Grammar,* (Ed, Haegeman, L.) Kluwer Academic Publishers, Dordrecht, pp. 198-235.
Miller, G.A. & Chomsky, N. (1963) Finitary models of language users. In *Handbook of Mathematical Psychology, vol. 2.,* (Eds, Luce, R.D., Bush, R.R. & Galanter, E.) Wiley, New York, pp. 419-491.
Newmeyer, F. (1980) *Linguistic Theory in America*. Academic Press, New York.
Newmeyer, F. (1986) *The Politics of Linguistics*. University of Chicago Press, Chicago.
Pearl, L. & Sprouse, J. (to appear) Computational models of acquisition for islands. In *Experimental Syntax and Island Effects,* (Eds, Sprouse, J. & Hornstein, N.) Cambridge University Press, Cambridge.
Perlmutter, D.M. (1983) *Studies in relational grammar*. University of Chicago Press, Chicago.
Phillips, C. (to appear) Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes*.
Pollard, C. & Sag, I.A. (1987) *Information-based Syntax and Semantics Volume 1: Fundamentals*. CSLI., Stanford.
Pollard, C. & Sag, I.A. (1994) *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.
Pollock, J.-Y. (1989) Verb movement, Universal Grammar and the Structure of IP. *Linguistic Inquiry*, **20**, 365-424.
Postal, P.M. (1971) *Crossover Phenomena*. Holt, Rinehart and Winston, New York.
Rizzi, L. (2004) *The Structure of CP and IP*. Oxford University Press, USA,
Ross, J.R. (1967) *Constraints on Variables in Syntax*. MIT.
Sag, I.A. (to appear) Sign-Based Construction Grammar: An informal synopsis. In *Sign-Based Construction Grammar,* (Eds, Boas, H.C. & Sag, I.A.) CSLI, Stanford, CA, pp. 39-170.
Sag, I.A., Hofmeister, P. & Snider, N. (2007) Processing Complexity in Subjacency Violations: The Complex Noun Phrase Constraint. *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*, Chicago.
Siddiqi, D. (2010) Distributed morphology. *Language and Linguistics Compass*, **4**, 524-542.
Staal, J.F. (1967) *Word order in Sanskrit and universal grammar*. Springer, Berlin.
Tomalin, M. (2006) *Linguistics and the formal sciences: the origins of generative grammar*. Cambridge University Press, Cambridge.
van Valin Jr., R. (2010) Role and Reference Grammar as a framework for linguistic analysis. In *The Oxford Handbook of Linguistic Analysis,* (Eds, Heine, B. & Narrog, H.) Oxford University Press, Oxford, pp. 703-738.
Van Valin, J., Robert D. & LaPolla, R.J. (1997) *Syntax. Structure, meaning and function*. Cambridge University Press, Cambridge.
Wasow, T. (1972) *Anaphoric Relations in English*. MIT.
Wasow, T. (1979) *Anaphora in Generative Grammar*. E. Story-Scientia, Gent.
Wasow, T. (2009) Remarks on grammatical weight. *Language variation and change*, **9**, 81-105.

Zwart, C.J.W. (1996) 'Shortest move' versus ‘fewest steps’. In *Minimal ideas: Syntactic studies in the minimalist framework,* (Eds, Abraham, W., Epstein, S., Thrainsson, H. & Zwart, C.J.W.) John Benjamins Publishing Company, Amsterdam: pp. 305-327.

NOTES

[1] I am grateful to Ray Jackendoff and Dan Siddiqi for comments on an earlier draft of this chapter and to the students and faculty of the 2012 Norwegian Summer School in Linguistics in Svolvaer for feedback and questions that have led to many significant improvements. All remaining errors are my responsibility

[2] For a survey of the work of the Sanskrit grammarians (around 1000 BC), see (Staal 1967). According to Staal, the Sanskrit grammarians were concerned with grammatical relations but not word order (Sanskrit being a free word order language). For a comprehensive history of more recent syntactic thinking, see Graffi 2001. For extended social, political and intellectual histories of generative grammar, see Newmeyer 1980; 1986; Matthews 1993; Tomalin 2006.

[3] For a more extended review, see Culicover and Jackendoff 2005: Chapters 1-Chapters 3.

[4] However, in the absence of a semantic theory in the 1960s, the distinction action/non-action had to be encoded syntactically. This was the approach taken by Generative Semantics (see §3.1), which assumed an abstract verb ACT only in (2a).

[5] With caveats for examples like *Sheila was sent flowers*. In this case, it is the indirect object that does not follow the verb.

[6] Although Chomsky (1981:212) continues to refer to A-over-A as a possible explanatory principle.

[7] In GS the structures in (22) were taken to be the DS representations of sentence (21), which was derived by transformation (in this case, lowering of the quantifier phrase).

[8] This is a simplification – it is assumed that both quantifier phrases undergo movement, and the scope is determined by which one is attached higher in the structure.

[9] Role and Reference Grammar (RRG) appears to be an exception; see §7.2.

[10] This statement is too strong, because an adverb can intervene between the verb and the direct object in English if the latter is “heavy” in some sense, e.g.,

(i) He ate quickly all of the fish on his plate.

For discussion of the factors that contribute to heaviness, see Wasow 2009.

[11] An extension of this approach appears in “cartographic” syntax, where the precise details of linear order are reduced to the hierarchical organization of functional categories both internal to the sentence and on the left periphery (Cinque 1999; Rizzi 2004).

[12] A number of devices have been proposed to achieve this result, e.g. unchecked syntactic features cause ill-formedness when mapped into PF and/or LF (Chomsky 1995).

[13] The movement of the subject does not pass through Spec,Agr_O since it is presumably filled by the direct object.

[14] The actual formulation of the LCA does not refer to a specific branching direction, but Kayne (1994: 33ff) argues that it reduces to precedence.

[15] Thus the MP bears a close resemblance to Categorial Grammar.

[16] For an extended review of phrase structure in generative grammar, see Carnie 2010, as well as Chapter 3.

[17] This is an oversimplification, since categories other than NP can be subjects, and subject in Russian (and other languages) may have other than NOMINATIVE case.

[18] These are different than the "functional" categories Infl, C^0, D^0 etc of GB. The GB categories should more accurately be termed "formal" categories, since they do not have to do with function. They have no meaning, but rather play a role in constraining the form of sentences.

[19] For discussion of the issue in Salish and Tagalog, see Koch and Matthewson 2009 and references cited there.