

226

ENCYCLOPEDIA OF COGNITIVE SCIENCE

2001

©Macmillan Reference Ltd

Syntax

syntax # constituent structure # transformations # Principles & Parameters # language acquisition

Phillips, Colin

Colin Phillips

University of Maryland USA

Syntactic theory aims to explain how people combine words to form sentences, and how children attain knowledge of sentence structure.

Goals of Syntactic Theory

Syntactic theory aims to provide an account of how people combine words to form sentences. A common feature of all human languages, both spoken and signed languages, is that speakers draw upon a finite set of memorized words and morphemes to create a potentially infinite set of sentences. This property of *discrete infinity* forms the basis of the creativity of human language, since it allows speakers to express and understand countless novel sentences which have never been uttered previously. Syntactic theory is concerned with what speakers know about how to form sentences, and how speakers come to attain that knowledge.

For example, speakers of English share the knowledge that *dogs chase cats* and *cats chase dogs* are possible sentences of English, but have different meanings. Speakers know that *chase dogs cats* is not a possible sentence of the language, and that *cats dogs chase* is possible in specific discourse contexts, as in *cats, dogs chase, but mice, they flee*. Speakers' knowledge of possible word combinations is often referred to as the (*mental*) *grammar*. The use of the term *grammar* here should not be confused with prescriptive grammars, which are socially imposed requirements on how sentences should and should not be composed (e.g. *don't split infinitives!*, or *say 'Sam and I' not 'me and Sam'*). Linguists use the term (*mental*) *grammar* to refer to

speakers tacit knowledge of how to create and interpret sentences, with no regard to considerations of social prestige or stigma.

An accurate model of a speaker's knowledge of his language should minimally be able to generate all and only the possible sentences of the language. For this reason, syntactic theory is often known as *generative grammar*. The roots of generative grammar go back to pioneering work by Noam Chomsky in the 1950s, which coincided with developments in computer science that made it feasible to explore explicit models of a speaker's syntactic knowledge. Attempts to create explicit generative grammars soon revealed that speakers' knowledge of syntax is a good deal more complex than had been anticipated. Research on syntactic theory has relied primarily upon speakers' intuitive judgments about the well-formedness (grammaticality) of sentences of their language. Since grammaticality judgments can be gathered quickly and easily from speakers of a wide variety of languages, syntactic theory has amassed a large database of findings about an ever more diverse set of languages.

The complexity of syntactic knowledge sharpens the problem of how language is learned. Furthermore, research on language acquisition has demonstrated that children know much of the grammar of their language before they are old enough to understand explicit instruction about grammar. Therefore, a primary challenge for syntactic theory has been to understand how a child can learn any language, relatively effortlessly, without explicit instruction.

Investigations of similarities and differences in the syntax of the world's languages take on greater importance in light of the language acquisition problem. *Universal* syntactic properties may be properties that the child does not need to learn, because they follow from the child's innate knowledge. Any non-universal syntactic properties that are found to cluster together across languages may also reduce the child's learning burden. The fact that the properties cluster together suggests that they all reflect a single deeper principle. Thus, the study of comparative syntax and the study of language learning are closely related.

The speed, accuracy, and uniformity of language learning have made the language acquisition problem a central concern of syntactic theory. To the extent that real-time language use (i.e., language production and comprehension) is similarly fast, accurate and uniform, it provides a similar challenge for syntactic theory: syntactic knowledge must have a form that can be deployed rapidly and accurately. In practice, however, issues of language use have exerted less influence on syntactic theory than have questions of language acquisition.

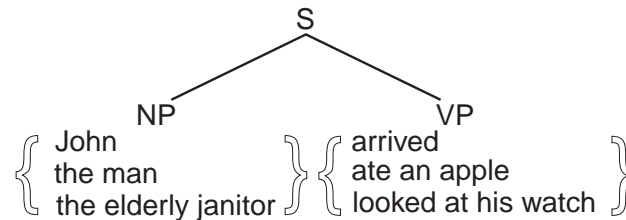
Fundamentals of Syntactic Theory

Discrete Infinity

Almost all accounts of the discrete infinity property of natural language syntax start from the notion that sentences consist of more than just sequences of words. In the minds of speakers and listeners, sentences are hierarchically *structured*

representations, in which words are grouped together to form phrases, which in turn combine with other phrases to form larger phrases. For example, a minimal sentence of English, such as *John arrived*, contains a subject and a predicate, but the roles of subject and predicate in the sentence may be replaced by phrases of arbitrary complexity. By representing possible subjects and predicates as *noun phrases* (NPs) and *verb phrases* (VPs) respectively, the structure of many possible sentences can be captured. This basic template for sentences of English can be expressed as a tree structure, as in (1a), or as a phrase structure rule, as in (1b). The two notations are interchangeable.

(1) a.



b. $S \rightarrow NP VP$

Just as rules like $S \rightarrow NP VP$ provide templates for sentences, templates can also be specified for the internal structure of noun phrases, verb phrases, and many other phrase-types. Even a small number of phrase structure rules and a small lexicon can generate large numbers of sentences. With only the 5 phrase structure rules in (2a) and a 9-word lexicon (consisting of 3 nouns, 3 determiners, and 3 verbs) 468 different sentences can be generated. If the lexicon is expanded to contain 10 words in each category, the grammar generates 122,100 different sentences. Such a system can express many different messages, but does not even begin to approach the creative power of human language.

(2)

S	\rightarrow	$NP VP$
VP	\rightarrow	$V NP$
VP	\rightarrow	V
NP	\rightarrow	$Det NP$
NP	\rightarrow	N

Rules which allow a phrase to be embedded inside another phrase of the same category are known as *recursive* rules. Coordination (3), modification (4), and sentential complementation (5) all involve recursion. They can thus be invoked arbitrarily many times in a single sentence. Such rules increase the expressive power of the grammar from merely vast to clearly infinite. There are obviously practical limitations on the length and complexity of naturally-occurring sentences, but such limitations are standardly attributed to general limitations on attention and memory.

(3)

NP	\rightarrow	$NP Conj NP$
VP	\rightarrow	$VP Conj VP$
$Conj$	\rightarrow	<i>and</i>

- (4) VP → VP PP
 NP → NP PP
- (5) VP → V S
 S → Comp S
- Comp → *that*

Clearly, the syntactic rules listed in (1-5) generate only a small fragment of English; many more rules need to be added in order to approach the expressive power of English. However, even this small fragment shows how natural language syntax is able to use finite means to generate infinitely many sentences.

Motivating Structures - Constituency

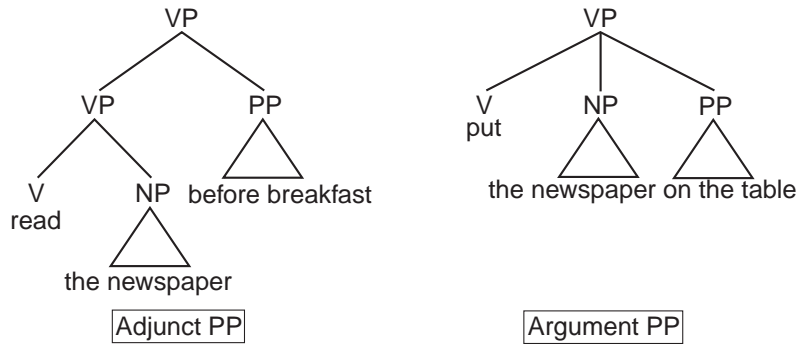
The syntactician's toolbox includes a number of structural tests which can be used as aids in diagnosing the structure of sentences. For example, constituents of sentences can generally be conjuncts in coordinate structures, as is shown for NPs and VPs in (6a-b). Other tests that show the constituency of VPs include substitution of the expression *do so* for a VP (7a), and fronting of the VP to a clause-initial position (7b).

- (6) a. Wallace fetched [_{NP} the cheese] and [_{NP} the crackers]
 b. Wallace [_{VP} sliced the cheese] and [_{VP} opened the crackers]
- (7) a. Wallace [_{VP} read the newspaper] and Gromit [_{VP} did so] too.
 b. Wallace wanted to [_{VP} impress Wendolene], and [_{VP} impress Wendolene] he did.

Constituency tests like those shown in (7) can be used to demonstrate that prepositional phrases (PPs) that are adjuncts (i.e., optional phrases) of VP recursively expand the VP, whereas PPs that are arguments of the verb do not. (8) shows that when *do so* substitution applies to a VP containing an adjunct-PP, the PP may be targeted or ignored by *do so* substitution. (9) shows that the adjunct-PP may be targeted or stranded by VP-fronting. This indicates that there is a smaller VP-constituent that excludes the adjunct-PP. In contrast, (10) and (11) show that an argument-PP cannot be ignored by *do so* substitution or VP-fronting. If the argument-PP is stranded by substitution (10b) or fronting (11b), the result is ungrammatical. This indicates that argument-PPs are contained within the smallest VP constituent. These contrasts motivate the VP-structures shown for adjuncts and arguments in (12).

- (8) a. Wallace [_{VP} [_{VP} read the newspaper] before breakfast],
 and Gromit [_{VP} [_{VP} did so]] too.
 b. Wallace [_{VP} [_{VP} read the newspaper] before breakfast],
 and Gromit [_{VP} [_{VP} did so] at lunchtime]

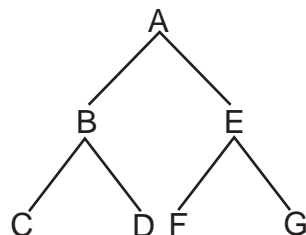
- (9) a. and [VP read the newspaper before breakfast] he did [VP ____]
 b. and [VP read the newspaper] he did [VP ____] before breakfast.
- (10) a. Wallace [VP put the newspaper on the table],
 and Gromit [VP did so] too.
 b. * Wallace [VP put the newspaper on the table],
 and Gromit did so on the floor.
- (11) a. and [VP put the newspaper on the table] he did [VP ____]
 b. * and put the newspaper he did on the table.
- (12) Adjunct PP: VP → VP PP
 Argument PP: VP → V (NP) PP



Motivating Structures — Hierarchy

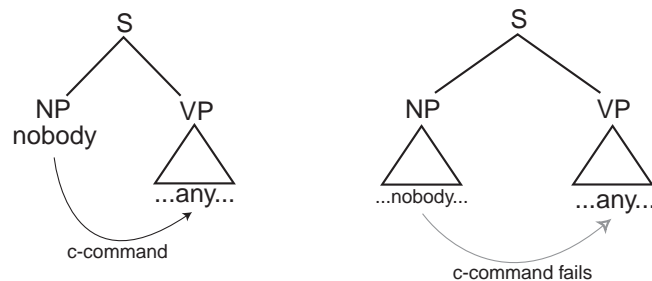
In addition to tools that show which groups of words form constituents, other tests diagnose the *hierarchical* relations among positions in a structure. A striking finding of syntactic research is that many grammatical phenomena are sensitive to a structural relationship known as *c-command*, which is similar to the logical notion of *scope*. A node *c-commands* its sister and any nodes contained inside its sister. Thus, in the structure in (13), node B *c-commands* node E, its sister, and nodes F and G contained inside its sister. Similarly, node E *c-commands* node B, its sister, and the daughters of B, nodes C and D. On the other hand, node C *c-commands* its sister, node D, and no others. Node A, the root node, does not *c-command* any nodes in the structure, since it has no sister node.

- (13) *C-Command*
 A node *c-commands* its sister and all nodes dominated by its sister.



One syntactic phenomenon that is sensitive to the c-command relation is the licensing of *Negative Polarity Items* (NPIs), such as *anybody* or *ever*. These items are only possible when they are c-commanded by an appropriate licenser, typically a negative expression like *not* or *nobody*. (14a-d) shows contexts in which NPIs are licensed; (14e-h) shows contexts where they are not licensed. The difference between the two sets of examples is that the negative expression c-commands the NPI in (14a-d), but fails to c-command the NPI in (14e-h), in which the NPI is embedded inside a subject NP.

- (14) a. Wallace didn't find any cheese.
 b. Nobody found any cheese.
 c. Wallace didn't think that he would ever refuse cheese.
 d. Nobody thought that Wallace would ever refuse cheese.
- e. * [NP The fact that Wallace didn't like the cheese] amazed anybody.
 f. * [NP The fact that nobody liked the cheese] amazed anybody.
 g. * [NP The person that Wallace didn't notice] thought that Gromit would ever return.
 h. * [NP The person that nobody noticed] thought that Gromit would ever return.



Similarly, in order for a pronoun to receive a bound-variable interpretation, the pronoun must be c-commanded by a quantificational NP. Thus, a bound-variable interpretation is available in (15a), but not in (15b), where the c-command requirement is not met. (15c-d) show that when the pronoun corefers with an NP that refers to an individual, the c-command requirement no longer applies. Following standard practice, variable binding and coreference are indicated in (15) by means of subscript indices.

- (15) a. Every boy_i loved the party that was thrown for him_i.
 b. * The party for every boy_i made him_i very happy.
 c. Harry_i loved the party that was thrown for him_i.
 d. The party for Harry_i made him_i very happy.

Multiple Roles: Transformations

In any sentence, each word or phrase occurs in a unique position in the linear sequence of words. However, many words and phrases appear to participate in

multiple structural relations in the sentence. A central concern of syntactic research since the 1950s has been to understand how speakers represent the multiple different roles that a single phrase can assume in a sentence.

The multi-functionality of phrases has been most fully explored in the case of NPs. A single NP can participate in a number of syntactic relations in a single sentence. First, speakers represent the *thematic role* of each NP in a sentence. *Agent* thematic roles are canonically realized on subject NPs, *theme* thematic roles are canonically realized on direct object NPs, and thematic roles like *goal*, *beneficiary*, or *location* are canonically realized inside PPs (16).

- (16) Wallace sent the flowers to Wendolene
agent theme goal

However, it is important to distinguish thematic roles from *grammatical relations* like *subject* and *object*, since thematic roles can be realized in different grammatical relations. The theme argument of the underlined verb *steal* is realized as a direct object in the active sentence in (17a), as a subject in the passive sentence in (17b), and as the subject of a higher clause in the raising construction in (17c).

- (17) a. The penguin stole the diamond. *active*
theme
 b. The diamond was stolen by the penguin. *passive*
theme
 c. The diamond seemed to have been stolen. *raising + passive*
theme

It is also important to distinguish thematic roles and grammatical relations from the structural positions in which NPs receive their *scopal* interpretation. In English, *wh*-phrases are positioned sentence-initially in direct *wh*-questions (18a), and in clause-initial position of an embedded clause in indirect *wh*-questions (18b). The *wh*-phrases in (18a-b) both receive a theme thematic role from the same verb, but receive different scopal interpretations. Other constructions in English in which NPs are displaced to their scopal position include topicalization (19a) and relative clauses (19b).

- (18) a. Which story did the teacher know that the children always like?
theme
direct object
matrix scope
 b. The teacher knew which story the children always like.
theme
direct object
embedded scope
- (19) a. This story, I think the children will like.
 b. The children hated the story which the teacher thought that they liked.

In English the scope of quantified NPs like *every boy* is not encoded in surface word order, with the consequence that sentences like (20) are scopally ambiguous. However, in languages like Hungarian and Japanese, the scope of quantificational NPs is often fixed by surface word order. (21a) has the canonical subject-before-object word order of Hungarian, and it is scopally unambiguous. The direct object can receive a wide-scope interpretation in the non-canonical word order in (21b).

- (20) Some student answered every question correctly.
some > every: a single student got all the questions correct
every > some: for each question, there was at least one student who got it right
- (21) a. Minden fi pontosan hat filmet l tott
 every boy exactly six films saw
 every > six: for every boy, there were exactly six films that he saw
 * *six > every: there were exactly six films that were seen by every boy*
- b. Pontosan hat filmet l tott minden fi
 exactly six films saw every boy
 every > six: for every boy, there were exactly six films that he saw
 six > every: there were exactly six films that were seen by every boy

Since speakers must distinguish for any NP a representation of its thematic role, its grammatical relation and its scope, it is important to ask how these different structural positions are related to one another.

The most well-known answer to this question is that speakers represent multiple structures for each sentence, one structure that encodes the thematic role assignments of a sentence, and one or more additional representations that encode grammatical relations and scope relations. The thematic representation is typically known as the *deep structure* (or *d-structure*) of the sentence, and it is related to the other levels of representation - *surface structure* (or *s-structure*) and *logical form* (LF) - by means of transformational processes which move NPs from thematic positions to positions associated with grammatical relations, and from there to scope positions. This *Transformational Grammar* approach was proposed by Chomsky in the 1950s and has been elaborated and modified by many linguists since that time.

(22) shows a sample derivation of a *wh*-question under this approach. The theme NP *who* receives the theme thematic role as the sister of the verb in the d-structure representation in (22a). In (22b) it moves to subject position of the embedded passive clause. Finally, in (22c) it moves from subject position to its clause initial scope position. Following a now-standard elaboration of the transformational approach, introduced in the 1970s, (22) represents the residue of movement as phonologically null elements known as *traces* (t).

- (22) a. the announcer believed had been elected who
 b. the announcer believed who_i had been elected t_i
 c. who_i did the announcer believe t_i had been elected t_i

Some alternative approaches share with transformational grammar the assumption that sentence structure representations consist of multiple levels, but deny that there is a derivational relationship among the levels of representation (e.g., Lexical Functional Grammar, Combinatory Categorical Grammar). Other approaches assume that the thematic, grammatical relation and scope properties of an NP are all simultaneously represented in a single enriched phrase structure, but that only one of these positions is normally encoded phonologically (e.g., Head-Driven Phrase Structure Grammar).

Types of Dependencies

Speakers' syntactic representations encode a number of different kinds of structural relations between words and phrases. The most basic relations are the groupings of words and phrases which form hierarchical constituent structures. These relations can be expressed entirely in terms of *sisterhood* and *dominance* among phrase structure nodes. The discussion above has introduced a number of other syntactic relations, or *dependencies*. Syntactic research has explored each of these dependencies in detail, and has investigated the extent to which apparently different dependencies follow a common set of principles.

The syntactic relation between a pronoun or reflexive and the NP from which it takes its reference is known as a *binding* dependency. In recent years, it has become increasingly common to distinguish true binding dependencies (23a), in which the *antecedent* of the pronoun or reflexive is a quantificational NP, from *coreference* dependencies (23b), in which the antecedent is a referring NP (e.g., *the dog*, *Wallace*, *a terrible argument*). Despite certain differences between the two dependency types (see (15) above), in English both types of dependencies are subject to the restriction that reflexives have a local (clausemate) antecedent, and that pronouns *not* have a local antecedent (24a-d) — such constraints have been studied extensively under the rubric of *Binding Theory*.

- | | | | |
|------|----|---|-------------------------|
| (23) | a. | Every Englishman _i thinks he _i is a great cook. | <i>variable binding</i> |
| | b. | Wallace _i thinks he _i is a great cook. | <i>coreference</i> |
| (24) | a. | Wallace _i likes himself _i | <i>local</i> |
| | b. | * Wallace _i thinks that Wendolene likes himself _i | <i>*non-local</i> |
| | c. | * Wallace _i likes him _i | <i>*local</i> |
| | d. | Wallace _i thinks that Wendolene likes him _i | <i>non-local</i> |

A number of other syntactic dependencies involve referential dependencies. In *VP-ellipsis* constructions the VP in the second conjunct is dependent on the VP in the first conjunct for its interpretation (25). Transformational analyses of *wh*-questions (26a) and relative clauses (26b) treat the relationship between the *wh*-phrase and the trace as a binding relation between a *wh*-operator and a variable.

- (25) Wallace [_{VP} likes cheese]_i and Gromit does [_{VP}]_i too.

- (26) a. Who_i did the voters elect t_i
 b. The man who_i the voters elected t_i

A leading question in research on referential dependencies involves how closely related the different types of referential dependencies are: does each type of dependency follow independent principles, or do they follow the same principles. There have been a number of interesting attempts to reduce movement and binding dependencies to a common set of underlying principles. However, these issues remain controversial.

Constraints on Dependencies

Wh-movement and related phenomena such as relativization and topicalization have been among the most extensively investigated topics in syntactic research, giving rise to a wealth of detailed findings about the properties of these constructions. By virtue of their length, syntacticians can easily manipulate which structural positions participate in the *wh*-dependency, and which structural positions the dependency crosses. Due to the semantic variety of the elements that may participate in these dependencies (e.g., *who*, *why*, *which of the candidates*, *John*, *a chicken*), there is substantial scope for manipulating the nature of the participants in the dependency. *Wh*-dependencies have thus served as a kind of magnifying glass for the investigation of syntactic dependencies.

Wh-dependencies can span many clauses, in fact arbitrarily many clauses, and thus they are often referred to as *unbounded dependencies*. In (27), the *wh*-phrase has been *extracted* from a number of embedded clauses, each of which is the *complement* (direct object) of the next higher verb.

- (27) Which candidate_i did the court fear [that the public might conclude [that the voters had elected t_i]]

However, in a tradition of research beginning with highly influential work in the late 1960s by John Robert Ross, it has been found that there are many syntactic environments which *wh*-extraction cannot cross. Following terminology introduced by Ross, the environments that block extraction are known as *islands*, and many of the restrictions on extraction are known as *island constraints*.

Relative clauses create islands for extraction (28a), as do indirect questions (28b), complements of NPs (28c), subjects (28d) and adjunct clauses (28e). Extraction from definite or specific NPs is highly marked (29b-c), although indefinite NPs create no such difficulties (29a). If a phrase is extracted from one conjunct of a coordinate structure, it must also be extracted from the other conjuncts (30a-b). Similar constraints apply to relativization and topicalization.

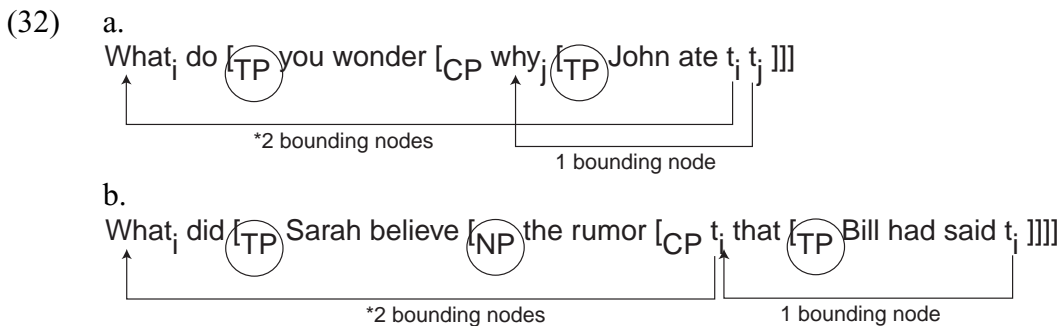
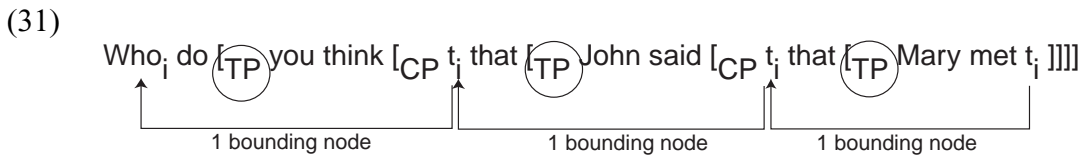
- (28) a. * Who_i did the court upset the voters [who favored t_i]
 b. * Who_i did Bill wonder [whether his new outfit would shock t_i]
 c. * What_i did Sarah believe [the rumor that Ed was willing to spend t_i]

- d. * Who_i did [the fact that the president nominated t_i] upset the opposition party?
 - e. * What_i did Wallace eat the cheese [while he was reading t_i]
- (29) a. Who_i did Sally hear a story about t_i ?
 b. ? Who_i did Sally hear the story about t_i ?
 c. * Who_i did Sally hear Helen s story about t_i ?
- (30) a. *What_i did [Gromit read the newspaper] and [Wallace eat t_i]
 b. What_i did [Gromit read t_i] and [Wallace eat t_i]

The examples in (27-30) raise the question of why *wh*-dependencies can be arbitrarily long in (27), but not in (28-30); put differently, why should relative clauses, *wh*-expressions, possessors, etc. interfere with the formation of *wh*-dependencies?

Work on this question led to the proposal that, contrary to appearances, all *wh*-dependencies are *local*. If all *wh*-movement is local, proceeding one clause at a time, then apparent instances of long-distance *wh*-movement turn out to be a series of local movements, each one targeting a landing site in the next higher CP (*complementizer phrase*), as in (31). Such sequences of local *wh*-dependencies are known as *successive cyclic movement*. If all *wh*-movement is forced to be local in this manner, then it follows that relative clauses and embedded *wh*-questions create islands, because these are cases in which one of the necessary intermediate landing sites of movement is already filled, as illustrated in (32a).

Probably the best-known implementation of this idea is the *Subjacency Constraint*, proposed by Chomsky in the early 1970s. In its original formulation the constraint blocks any *wh*-dependency which spans more than two *bounding nodes*, where the bounding nodes are defined as NP and TP (*Tense Phrase*: a more recent term for the S node, which recognizes *Tense* as the head of a clause). This formulation also explains the complex NP constraint violation in (28c) and (32b).



The proposal that a long-distance *wh*-dependency involves a sequence of local dependencies receives interesting support from a number of languages which show a syntactic residue of successive cyclic movement. In certain varieties of Spanish, subject-auxiliary inversion occurs in every clause in the path of *wh*-movement (33: compare this to the English translation, in which inversion occurs only in the highest clause). In varieties of English spoken in certain parts of Northern Ireland the *wh*-expression *who all* can be split, stranding the word *all* in the embedded CP (34).

- (33) a. Juan pensaba que Pedro le hab a dicho que la revista hab a publicado ya el articulo.
 J. thought that Pedro him had told that the journal had published already the article
 Juan thought that Pedro had told him that the journal had published the article already.
- b. Qu pensaba Juan que le hab a dicho Pedro que hab a publicado la revista?
 What thought Juan that him had told Pedro that had published the journal
 What did Juan think that Pedro had told him that the journal had published?
- c. * Qu pensaba Juan que Pedro le hab a dicho que la revista hab a publicado?
 What thought Juan that Pedro him had told that the journal had published
- (34) *West Ulster English*
- a. What all do you think (that) he ll say (that) we should buy t?
 b. What do you think all (that) he ll say (that) we should buy t?
 c. What do you think (that) he ll say all (that) we should buy t?

The island constraints restrict the nodes that a *wh*-dependency may cross. All of the examples presented so far involve extraction of a direct object *wh*-phrase. In addition, subject and adjunct *wh*-phrases in English are subject to tighter restrictions than object *wh*-phrases. For example, extraction of an embedded direct object *wh*-phrase is possible, irrespective of whether the embedded clause contains an overt complementizer *that* (35a-b). However, extraction of an embedded subject *wh*-phrase is impossible if the complementizer is overt (35c). This constraint is known as the *that-trace constraint*, and it has been observed in many languages, as discussed further below.

- (35) a. Who_i do you think t_i that John met t_i ?
 b. Who_i do you think t_i John met t_i ?
 c. * Who_i do you think t_i that t_i met John ?
 d. Who_i do you think t_i t_i met John ?

There are also differences in extraction possibilities between argument *wh*-phrases like *what* and *which books*, and adjunct *wh*-phrases like *why* and *how*. A direct object *wh*-phrase can be extracted from an infinitival indirect question (36), but an adjunct *wh*-phrase cannot, as shown by the unavailability of an embedded interpretation of *why* in (37a). (37b) shows that the embedded interpretation of *why* is available when *how* does not intervene.

- (36) What did John say how to repair ___?
- (37) a. Why did John say how to repair the car?
matrix clause why ok, embedded clause why impossible
- b. Why did John say that he should repair the car?
matrix and embedded why both ok

A long-standing goal of syntactic research on unbounded dependencies has been to uncover a set of general principles which can explain the full variety of constraints on *wh*-dependencies. Although there have been many different attempts to unify the constraints on movement, two observations have been pervasive, and have featured in many different theories. First, if movement is required to be *local*, then it is subject to *intervention effects*, when a required landing site of movement is occupied by another element. Second, movement paths which include non-complement nodes (subjects or adjuncts) are consistently more restricted than paths which include only complement nodes (sisters of heads).

Cross-Language Similarities & Differences

Cross-linguistic studies have played a major role in the development of syntactic theories. A fully general theory of the mental representation of syntax clearly must handle the facts of all human languages. In addition, cross-linguistic investigations are important to accounts of how natural language syntax is learnable. *Universals* of syntax, or *principles*, may be part of the child's innate endowment, and thus not need to be learned. *Non-universal* syntactic properties must also be learnable within the constraints imposed by the time and evidence available to the child. When a set of syntactic properties *covaries* across languages, it is possible that the learner only needs to learn one of the set of properties in order to draw appropriate conclusions about the entire set of properties. Thus, an important goal of cross-linguistic syntax research is to find clusters of covarying syntactic properties, or *parameters*. This formulation of the language acquisition problem is known as the *Principles and Parameters* (P&P) approach. The P&P approach has been most intensively investigated in transformational approaches to syntax, since around 1980, but it can be applied equally well to other syntactic approaches and to other areas of linguistics.

Research on comparative syntax has discovered a number of striking cross-linguistic parallels between languages that appear very different on the surface. An example from Mohawk serves as an illustration.

One of the constraints on pronouns in English is that a pronoun cannot corefer with an NP that it c-commands. This constraint (Binding Condition C) explains the

contrast between (38a) and (38b). By virtue of the configurational structure of English clauses, the pronoun inside the subject NP fails to c-command the direct object in (38a), thereby allowing coreference. On the other hand, the subject pronoun c-commands the pronoun inside the object NP in (38b), thereby blocking coreference.

- (38) a. [NP The book that he_i bought] offended John_i
 b. * He_i bought [NP the book that offended John_i]

In a series of detailed studies of Mohawk, an Iroquoian language spoken in Quebec and upstate New York, Mark Baker has shown that Mohawk shows similar configurational properties to English, despite initial appearances to the contrary. Unlike English, which exhibits strict subject-verb-object (SVO) word order, Mohawk exhibits free word order, allowing all six possible permutations of subject, verb and object (39). Mohawk allows omission of subjects and objects, compensating with a rich system of agreement morphology (40), and allows discontinuous constituents in the form of split noun phrases (41). Based on properties like these, languages like Mohawk have often been described as non-configurational, suggesting a lack of configurational contrasts like the one shown for English in (38). However, Baker shows that in multi-clause sentences, Mohawk exhibits similar configurational asymmetries to English, as shown by the contrast in coreference possibilities in (42). This contrast can be explained by Binding Condition C, in the same way that the contrast in (38) was explained in English. This in turn requires that we assume that Mohawk has a configurational structure, similar to English.

- (39) a. Sak ra-n hwe -s ako-[a]ty tawi
 Sak MsS-like-HAB FsP-dress
 Sak likes her dress.
- b. Ra-n hwe -s Sak ako-[a]ty tawi
 c. Sak ako-[a]ty tawi ra-n hwe -s
 d. Ra-n hwe -s ako-[a]ty tawi Sak
 e. Ako-[a]ty tawi ra-n hwe -s Sak
 f. Ako-[a]ty tawi Sak ra-n hwe -s

- (40) Ra-n hwe -s
 MsS-like-HAB
 He likes it.

- (41) Ne k ke wa-hi-y na- ne kw skwes
 NE this FACT-1sS/MsO-catch-PUNC NE pig
 I caught this pig.

- (42) a. Wa-ho-nakuni- tsi Sak wa-hi-hrewaht-e
 fact-NsS/MsO-anger-punc that Sak fact-1sS/MsO-punish-punc
 That I punished Sak_i made him_i mad. (coreference possible)
 b. Wa-shako-hrori- tsi Sak wa-hi-hrewaht-e
 fact-MsS/FsO-tell-punc that Sak fact-1sS/MsO-punish-punc
 He_i told her that I punished Sak_i. (coreference impossible)

The similarity between coreference possibilities in English and Mohawk is striking, given how different the two languages appear on the surface. Furthermore, similar constraints on coreference have been observed in countless languages. Therefore, it is reasonable to conclude that Binding Condition C is a universal of natural language syntax, which may be part of innate linguistic knowledge. Consistent with this suggestion, studies by Stephen Crain and his colleagues have shown that children exhibit knowledge of Binding Condition C at least as early as their third birthday, which is as early as it has been possible to test this knowledge. This finding has been replicated in children learning other languages (e.g., Italian, Dutch, Russian). The conclusion that Binding Condition C need not be learned is particularly good news for a child learner of Mohawk, since it is unlikely that the presence of the Condition C constraint could be inferred from the input to the Mohawk child.

Although syntactic research has uncovered many universals of language, there are clearly many properties that vary across languages and hence must be learned. The search for parametric clusters of syntactic properties has turned up a number of cross-language correlations. From the perspective of learnability, the most useful correlations are those which link abstract (and hence difficult-to-observe) syntactic properties with more easily observable syntactic properties. For example, the *that-trace* constraint introduced above does not apply in all languages: it applies in English (43a), but not in Italian (43b). This is a contrast that is not easily inferred from the language input to children, since it is not easy to observe the *absence* of a particular construction. However, the availability of *that-trace* sequences correlates cross-linguistically with the availability of post-verbal subjects, which are readily observable in the input to the learner. Italian allows post-verbal subjects (44b), but English does not (44a). This connection has been reinforced based on the study of dialects of Arabic from the area of Jordan and Palestine. Whereas the Beni-Hassan dialect allows post-verbal subjects and *that-trace* violations, the nearby Levantine dialect allows neither post-verbal subjects nor *that-trace* violations. Therefore, the child learner may be able to learn whether the *that-trace* constraint applies in his language, by observing whether post-verbal subjects are available.

- (43) a. * Who_i did you say that t_i has written this book?
 b. Chi_i hai detto che t_i ha scritto questo libro?
 who have-you said that has written this book
 Who did you say has written this book?
- (44) a. * Have arrived many students
 b. Hanno arrivato molti studenti.
 have-3pl arrived many students
 Many students have arrived.

A number of other cross-language correlations have been discovered, which link hard-to-observe syntactic properties to more easily observable phenomena. However, this line of research is still in its infancy. A good deal of work remains to be done, before it can be shown that all cross-linguistic variation in syntax can be reduced to a set of easily learnable parametric clusters.

Variants of Syntactic Theory

Over the past 40 years, syntactic theory has undergone a number of changes, and has spawned a variety of different grammatical theories, each with a different title, such as Relational Grammar (RG), Head-Driven Phrase Structure Grammar (HPSG), Lexical-Functional Grammar (LFG), Categorical Grammar (CG), Government-Binding Theory (GB), Tree Adjoining Grammar (TAG), etc. While it is tempting to view these as monolithic alternatives, to do so would be misleading.

First, all approaches provide only fragments of a full theory of grammatical knowledge; sometimes these fragments only partially overlap between approaches. For example, the emergence of Relational Grammar in the 1970s led to a wealth of insights about grammatical relation changing processes (such as passive, raising, etc.), but had little to say about the representation of scope or constituency. The insights of this theory have been incorporated into many other syntactic frameworks. Second, there are many fundamental points of agreement between the different approaches, and in some respects opinions are converging more than they are diverging. Finally, the differences among practitioners of the same general framework can be as large or even larger than the differences between different approaches. The differences of opinion that engender different named grammatical theories naturally draw greater attention, but they have no special status. Therefore, rather than reviewing different named grammatical theories, this section focuses on a selection of fundamental issues on which syntactic theories diverge.

Syntactic Atoms and How they Combine

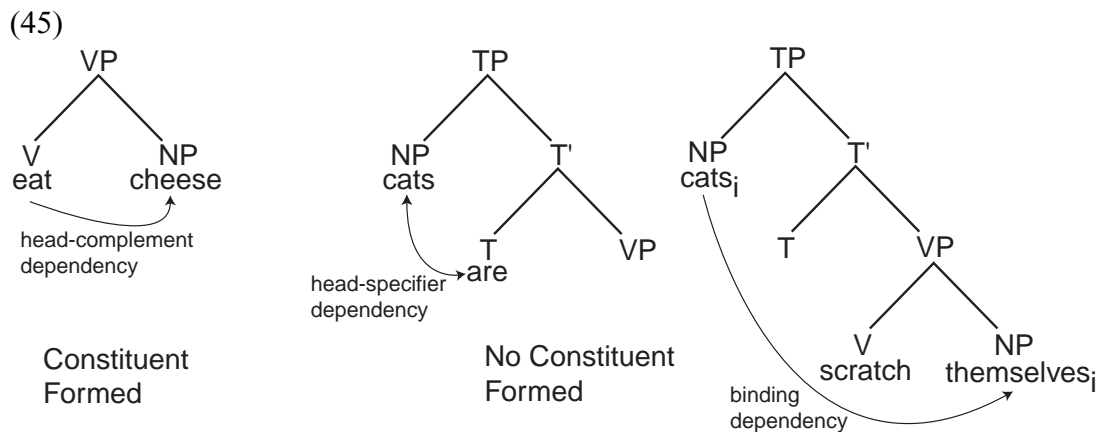
First, syntactic theories differ on the issue of what are the atoms of syntax, i.e. the pieces of sentences that are stored in a speaker's long-term memory, and that are combined to form syntactic structures. At one extreme are certain versions of Transformational Grammar (including the recent *Minimalist Program*), which claim that the atoms of syntax are smaller than words — either morphemes or individual syntactic features. Under this approach, underlying syntactic structures are formed by selecting a set of these atomic units, and combining them based on highly general principles of structure-building. Under this approach, syntax is responsible even for the formation of word-sized units. For example, an inflected verb like *runs* may be formed by independently selecting the verbal head *run* and the inflectional head [*3rd person singular, present*], and applying a transformation which combines them to form a complex syntactic head, which is spelled out as the word *runs*.

At the other end are approaches which assume much larger atoms, in the form of templates for phrases or even clauses. Construction Grammar and some versions of Tree Adjoining Grammar (TAG) are examples of such approaches. Under these approaches, the representation of idiomatic expressions is little different from the representation of other types of phrases. Construction Grammar has provided some insightful analyses of constructions that have been largely overlooked in mainstream transformational syntax. A promising area for future research is a comparison of the relative merits of theories with larger and smaller syntactic atoms.

Despite disagreements about the size of the atoms of syntax, the field of syntax has undergone a quiet convergence of opinion on the role of the atoms of syntax. In early generative syntactic theories it was standard to distinguish the terminal elements of syntax (i.e., lexical items), from the phrase structure rules which determine how the terminal elements combine. In most current syntactic theories this distinction has been eliminated, and the work once done by phrase structure rules is replaced by a set of highly general conditions on how syntactic atoms combine. In these *lexicalized* grammars, information about the combinatorial possibilities of syntactic atoms is built into the lexical entries of the atoms themselves. Lexicalism is a common feature both of theories which assume very small syntactic atoms and theories which assume much larger syntactic atoms.

Types of Structural Dependencies

A second issue involves the question of how syntactic elements enter into structural dependencies. It is useful to take as a starting point the phrase structure grammar notions assumed above. In a typical phrase structure grammar syntactic elements enter into two basic types of dependencies with other elements, illustrated in (45). First, when syntactic elements enter into a *sisterhood* relation with another element, this both forms a dependency between the two elements and creates a new syntactic *constituent*. Constituents may participate a variety of different syntactic processes, such as coordination, movement and ellipsis. On the other hand, many syntactic dependencies do not involve the formation of new constituents. For example, subject-verb agreement involves a dependency between two elements which do not form a syntactic constituent. Similarly, dependencies formed by reflexive binding connect a pair of elements which do not form a constituent.



The phrase structure notions which create this bifurcation among structural dependencies continue to dominate thinking about syntax, but there are a number of interesting alternative proposals which reduce or eliminate this distinction. First, *dependency grammars* treat all syntactic dependencies in a parallel fashion, and do not single out constituent-forming dependencies between sisters as special (46).

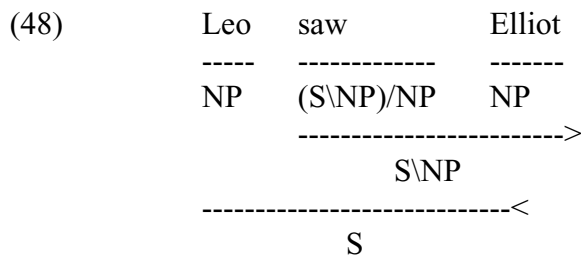
(46) *Dependency Grammar representation of argument & agreement relations*



Combinatory Categorical Grammars (CCG) also reduce the bifurcation of dependency types found in phrase structure grammars, but in the opposite manner from Dependency Grammars. In CCG grammars, information about the elements that a syntactic atom may combine with is encoded in enriched category labels. For example, an intransitive verb like *run* might have the category label $S \backslash NP$, which is read as a category which combines with an NP to its left to form an S. Dependencies between syntactic sisters are formed by the rule of *function application* (49). Dependencies between non-sisters are also formed by function application, thanks to the mediating effects of *function composition* rules (49a), which allow the combinatorial requirements of a syntactic atom to be passed up through a series of larger units.

By virtue of their more uniform treatment of structural dependencies, both Dependency Grammar and Combinatorial Categorical Grammars have spawned innovative proposals about the treatment of constituency phenomena. Whereas most phrase structure grammars impose a clear boundary on which syntactic relations form constituents, DG and CCG do not. Therefore, these approaches have been used to analyze syntactic phenomena which do not fall straightforwardly under standard notions of constituency, such as non-constituent coordination involving subject-verb sequences, and intonational phrases which do not obviously correspond to constituents in more standard approaches. In CCG, flexible constituency relations are made possible by the introduction of *type raising* rules, which modify the category label of a syntactic element in such a way that its combinatorial requirements can be satisfied in a different order (51-52).

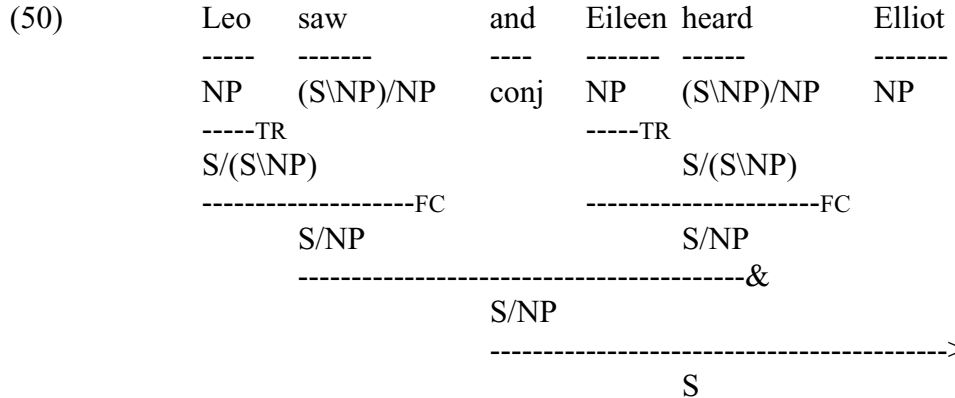
- (47) a. *Forward Function Application* ($>$)
 $X/Y \ Y \Rightarrow \ X$
 b. *Backward Function Application* ($<$)
 $Y \ X \backslash Y \Rightarrow \ X$



- (49) a. *Forward Function Composition* (FC)
 $X/Y \quad Y/Z \Rightarrow \ X/Z$ *general form*
 $S / (S \backslash NP) \quad (S \backslash NP) / NP \Rightarrow \ S / NP$ *specific case*

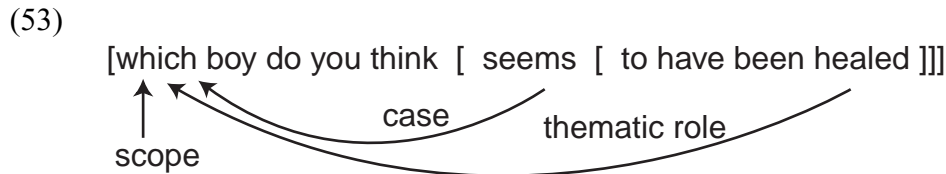
b. *Type Raising (TR)*
 $X \Rightarrow Y/(Y \setminus X)$ *general form*
 $NP \Rightarrow S/(S \setminus NP)$ *specific case*

c. *Coordination (&)*
 $X \text{ CONJ } X \Rightarrow X$



Multiple Roles: Alternatives to Transformations

All syntactic theories must address the fact that individual syntactic elements can enter into a variety of different structural relations. In a multi-clause sentence, a single phrase may take scope in one clause, be case-marked by the predicate of a second clause, and receive a thematic role from a verb in a third clause, as in (53).



Since the 1950s and 1960s, transformational approaches to syntax have famously argued that phrases can bear multiple syntactic roles because there are multiple syntactic levels of representation, which are related to one another via movement of words and phrases through a series of different structural positions. Typically, only one of these positions is reflected in the surface form of the sentence. As a result, one of the main areas of research on transformational syntax has been the investigation of constraints on possible movement operations, as outlined above.

Transformational analyses have attracted a great deal of attention both within linguistics and beyond. This is partly due to the strong claims that transformational analyses imply about the abstractness of syntactic representations, which in turn imply correspondingly strong claims about the innateness of linguistic knowledge. It is also partly due to the early claim that the underlying deep structure reflects the meaning of a sentence. This claim is no longer endorsed as strongly as it once was, although it crops up on a regular basis in various guises. Transformational approaches to syntax have become the most widely taught and researched approaches, and hence they have

also been the starting point for most alternative approaches, and the target of some of the most strident criticism.

The syntactic frameworks *Lexical Functional Grammar* (LFG) and certain versions of *Combinatory Categorical Grammar* (CCG) share with transformational grammar the assumption that words and phrases bear multiple roles because they appear in multiple different levels of representation. These approaches differ from transformational grammar in the respect that they do not assume that each different level is a hierarchical constituent structure, or that the levels are related by movement transformations. LFG assumes independent levels of *a-structure* (argument structure: representation of argument/thematic roles), *f-structure* (function structure: representation of subject, object etc. roles), and *c-structure* (constituent structure: surface syntax). There are rules for mapping between these levels, but the mappings are not assumed to be transformational. In some versions of CCG separate representations of *argument structure* and *surface structure* are posited.

In *Head-Drive Phrase Structure Grammar* (HPSG) only one syntactic level of representation is assumed. This single level of representation combines words and phrases into the kind of surface constituent structure that is familiar from many other syntactic theories. However, the terminal elements of these structures contain highly articulated feature structures, which encode a great deal of information about argument structure, phonology, moved arguments, etc. Whereas transformational grammars use movement operations to handle the multiple roles problem, in HPSG the same work is done largely internal to individual syntactic heads. For example, a verbal head may encode the information that a *wh*-phrase, which is represented as the *focus* argument of the clause, is to be treated as the filler of one of the slots in the verb's argument structure list. Constraints on movement operations in transformational approaches must be replaced in non-transformational theories by related constraints on the relations between scope and argument slots.

Causes of Ungrammaticality

Standard approaches to syntactic theory assume a set of syntactic atoms and a relatively small number of formal principles or constraints that determine how these atoms may be combined to form sentences. A sentence is assumed to be grammatical if it violates none of the formal constraints. An ungrammatical sentence is a sentence that violates one or more constraints. Some variants of syntactic theory have explored broader notions of what causes a sentence to be (un)grammatical.

Functional grammars typically emphasize the role of meaning or of communicative efficiency in determining the well-formedness of a sentence. Such approaches typically do not appeal directly to semantics or processing efficiency to explain ungrammaticality. Rather, semantics or processing efficiency are used to provide a functional motivation for a set of formal grammatical constraints.

Both *Optimality Theory* and certain versions of the *Minimalist Program* challenge the standard assumption that a grammatical sentence is a sentence that violates no constraints. This characterization is replaced in these approaches with the requirement that a well-formed sentence is the *optimal* candidate from a set of

possible structures/derivations for that sentence. In other words, a sentence may be deemed ungrammatical for the simple reason that there exists a better way of expressing the same thing.

Challenges & Future Prospects

Universal Grammar

Syntacticians are impressed by the fact that people show complex and subtle knowledge of their native language, which extends far beyond anything that they are taught or are consciously aware of. A primary goal of syntactic theory is to explain how any child can attain this level of knowledge, for any language of the world, with little apparent effort. The principles and parameters approach to comparative syntax, which is compatible with any of the syntactic frameworks discussed above, contributes to achieving this goal by seeking universal syntactic principles and clusters of syntactic properties which covary across languages. This approach provides a clear program of research, and a clear criterion for evaluating the progress of the program. Has syntactic variation yet been reduced to a set of universal principles and a small number of parametric clusters?

Syntactic research in the past 30 years has certainly uncovered a wealth of cross-linguistic findings. A number of good candidates for universals of syntax have been found. However, the search for parameters has met with mixed success. On the negative side, parametric clusters of properties appear to be both narrower and more numerous than originally expected. On the positive side, the prospect that each parametric cluster may be linked to an easily observable surface property of the language appears to be viable. A great deal of work remains to be done to fulfill the goals of the Principles and Parameters research program. It remains to be seen whether all of natural language syntax, including all of the idiosyncrasies of each language, can be handled in these terms.

Semantics & Phonology

In early work on syntax, relatively little attention was given to the relation between syntactic representations and semantic or phonological representations. In recent years there has been a substantial growth in the number of studies of the *interfaces* between syntax and other components of linguistic knowledge, particularly the syntax-semantics interface, and these areas will continue to be important in coming years.

The Unification Problem

Syntactic theories are mentalistic theories. Syntactic research aims to characterize what a person knows when that person knows a language, and how the person came to attain that state of knowledge. However, most syntactic theories characterize knowledge of which sentences are grammatical and which sentences are ungrammatical, with no suggestions about how speakers successfully access this

knowledge in real time (e.g., in speaking or understanding), or about how children come to acquire this knowledge. Even less is known about how to encode this knowledge in brain structures. The overall goals of syntactic theory may be affected by findings in these areas.

In the area of language acquisition, a variety of valuable experimental techniques have made it possible to discover what children know about syntax at specific ages, in children as young as one year of age. In many instances, surprising sophistication has been discovered in very young children. However, there is still very little understanding of how input to children drives *change* in their syntactic knowledge, and findings in this area may have broad implications. Syntactic research generally credits the learner with a severely limited ability to extract information from the input, and hence syntactic theory assumes the burden of finding extremely simple *cues* for all non-universal syntactic properties. Empirical studies of what input information learners are and are not able to use has the potential to either increase or reduce the burden on syntactic theory to narrow the scope of what must be learned.

In the area of real-time language comprehension and production, syntactic theories typically provide an account of what structures are ultimately built, but little information about the explicit procedures that are used to create these structures in real-time. Traditionally, such questions have been taken to be peripheral implementational issues, which have little impact upon the form of syntactic knowledge. However, a number of recent proposals question the standard distinction between real-time structure-building procedures and a more abstract competence grammar, and argue that syntactic knowledge can be better understood if it is viewed as an explicit procedural system. If true, then the well-known debates about derivational vs. non-derivational approaches to syntax become moot, since syntactic structure-building in comprehension and production is obviously incremental and derivational.

The ultimate questions of how syntactic knowledge is encoded in the brain and in the genome remain largely mysterious. Much has been learned about which regions of the brain are important to syntactic processing, and it has become possible to pinpoint specific patterns of neural activity which reflect syntactic processing. However, such investigations have led us no closer to an understanding of how groups of neurons *encode* syntactic information, in such a way as to solve the discrete infinity problem at the neuronal level. In the short term, at least, advances in this area are most likely to come from work in theoretical and computational neuroscience, rather than from recordings of human brains. The problem in this area is that it is not yet clear what to look for.

Further Reading

Baker, M (1996) *The Polysynthesis Parameter*. New York: Oxford University Press.

Baltin, M & Collins C (2000) *Handbook of Contemporary Syntactic Theory*. Malden, MA: Blackwell.

- Bresnan, J (2000) *Lexical Functional Grammar*. Malden, MA: Blackwell.
- Chomsky, N (1995) *The Minimalist Program*. Cambridge, MA: MIT Press.
- Culicover, P (1997) *Principles and Parameters*. New York: Oxford University Press.
- Radford, A (1988) *Transformational Grammar*. Cambridge, UK: Cambridge University Press.
- Roberts, I (1997) *Comparative Syntax*. London: Edward Arnold.
- Sag, I & Wasow, T (2000) *Syntactic Theory: A Formal Introduction*. Stanford, CA: CSLI Publications.
- Steedman, M (2000) *The Syntactic Process*. Cambridge, MA: MIT Press.