The Theoretical Orientation of *The Cambridge Grammar of the English Language*

Rodney Huddleston and Geoffrey K. Pullum

The long tradition of English grammatography stretches back to the late 16th century, and was informed by a classical tradition much older than that. The achievements of the early grammarians are certainly something to marvel at. The pioneer, William Bullokar (1586), navigating solely by the unreliable star of Latin, posited five cases for English nouns despite the absence of any case inflection, but by the following century John Wallis's grammar *Grammatica Linguae Anglicanae* (1653), though written in Latin, explicitly rejected the notion that English nouns had grammatical case or gender (Linn 2006, 74–75).

By 1762, when Robert Lowth published *A Short Introduction to English Grammar*, the idea that English was a disreputable language whose scruffiness needed to be concealed within Latin vestments had largely faded. Lowth, rather unfairly portrayed today as the father of obdurate and unmotivated prescriptivism (Pullum 1974), was well aware that English has preposition stranding whereas Latin does not. He called it "an Idiom which our language is strongly inclined to" — deliberately using the construction himself (humourless plagiarizers later rephrased the remark as "an idiom to which..."; see Tieken-Boon 2011, 115–116). He also understood its status as relatively informal style: "it prevails in common conversation" and in "the familiar style in writing".

However, the evolution of grammatical analysis of English slowed to a crawl after Lowth's time, and eventually almost stopped. Works produced for school students and the general public hardly changed their accounts of elementary matters like the definitions of the 'parts of speech' or the classification of subordinate clauses in the following 250 years. (The rise of structural and generative theoretical linguistics had essentially no influence at all on the teaching of grammar in schools, or on material addressed to the general public.) English grammar was treated as a body of dogma to be revered, obeyed, and promulgated — not as a topic for evidence-gathering or investigation. Virtually every work aimed at school students or the general public over several centuries repeated the traditional dogma uncritically in essentially the same form. Little more than style differentiates the statements made in books published in 2000 from books published in 1900 or the early 1800s.

Our admiration for the accomplishments of scholars like Bishop Lowth should not imply that his analysis should continue to be accepted without revision and presented to schoolchildren and general readers today. Yet this is broadly what happened.

"The PREPOSITION", says Lowth (1762), is "put before nouns and pronouns chiefly, to connect them with other words, and to show their relation to those words." "PREPOSITIONS serve to connect words with one another, and to show the relation between them," says Lindley Murray (1795), closely tracking Lowth. "A **preposition** is a word used to show the relation between its object and some other word," says Thomas Harvey six decades later (1868). "A Preposition ... shows in what relation one thing stands to another thing," says Nesfield (1900) at the turn of the 20th century. "A *preposition* is a word which governs a

noun or a pronoun and connects it to anything else in the sentence or clause," says Gwynne (2011) after another hundred years and more has gone by. Grammar books are simply reiterating what they take to be ancient wisdom, paraphrasing whatever the last one said. They are not engaging critically in the investigation of syntactic structure. (As we remark later, the quoted statements about prepositions, taken as serious attempts at a definition, are utterly indefensible.)

The Cambridge Grammar of the English Language (Huddleston & Pullum 2002, henceforth *CGEL*) takes the view that it is not acceptable to preserve misguided grammatical concepts or analyses simply out of reverence for the grammarians of past centuries. Intended primarily as a reference grammar for scholars with a professional interest in the structure of contemporary Standard English, *CGEL* sticks with traditional terminologies and assumptions wherever that is reasonable (there is no virtue in neologism simply for its own sake), but cuts ties with the tradition wherever it is conceptually unintelligible or empirically indefensible. Without presupposing a technical training in linguistics, it also attempts to incorporate insights from compendious grammars like Jespersen's classic *A Modern English Grammar on Historical Principles* (1909-1949); structuralist works like Bloomfield's *Language* (1933); the data-centred research of the Survey of English Usage that culminated in the Quirk team's magnum opus *A Comprehensive Grammar of the English Language* (1985; see Huddleston 1988 for a review); and thousands of generative grammatical studies over the past six decades. This chapter surveys some of the key arguments that motivate *CGEL*'s revisions and emendations of the tradition.

Category and function

The deepest problems with traditional grammar stem from its tacit assumption that grammatical categories can be defined in terms of vaguely delineated word meanings. Lurking behind this assumption is a deep confusion about the difference between the classification of words into classes or categories and the identification of what role or function a word is serving within a particular construction. We begin with a discussion of this issue, since a clear and sharp distinction between category and function plays a major role in *CGEL*'s analysis.

A **category** is a collection of words or phrases that share certain grammatical properties: 'noun' (N) and 'noun phrase' (NP), for example. A word's dictionary entry will include information about the category (or categories) to which it belongs. And phrases, too, are assigned to categories like NP on the basis of their form, regardless of the structure of the surrounding sentence.

The **function** of a syntactic unit is the grammatical relation it bears to the larger construction containing it, or to another element within that construction. In *Some people closed their windows*, for example, *some people* and *their windows* belong to the same category, NP, but they have different functions—different relations to the clause or to the verb *closed*: they are respectively the Subject and the Object. (We adopt the convention of using initial capitals for the names of functions like Subject, Object, Head, Complement, Modifier, Coordinate, etc., and not for category names like 'noun phrase' or 'adjective' or 'clause' — though of course abbreviations like 'NP' are also standardly written in capitals.)

Dictionaries can never give information about functions in this sense, because the function of an item is not intrinsic to it, but rather relational—it is dependent on the structure of the sentence in which it appears. Thus while dictionaries can and do indicate that *pork* is a noun, they cannot identify *pork* as a Subject: in *Pork is delicious* it is, but in *I like pork* it isn't. They cannot say whether *pork* is a Coordinate (i.e., one of the coequal members of a coordination), because sometimes it is (as in *How about pork and beans?*) and sometimes it isn't (as in *Do you like pork?*).

We will return to the distinction between category and function and make crucial use of it at several points in what follows.

The mistake that traditional grammar books make in their definitions of lexical categories is to attempt to give definitions on what is in essence a universally-oriented basis (though they do not generally acknowledge this). Thus the definition of 'noun' will be one that enables us (at least very broadly) to see why 'noun' is used not just when talking about certain English words but also about certain words with comparable meanings in Japanese and Swahili and thousands of other languages. Giving a universal characterization of such a term is a task to be carefully distinguished from that of providing necessary and sufficient conditions for categorizing words within a language. Traditional grammars do not even attempt to draw this distinction.

To define a notion like 'noun', traditional grammars rely on vague intuitions about meaning: they invariably define nouns as words that name things. It is indeed true that the words for naming temporally stable entities and physical material are included among the nouns, in any language, but that cannot be the basis for a definition. The absurdity of any such basis is not sufficiently recognized. The assumption implicit in the traditional definition is that we can identify 'things' independently of the words used to denote them and then define nouns as the words that denote these things. It implies that we can ascertain without reference to language that there are such things as clocks, clouds, cuckoos, colours, chances, correlations, costs, carelessness, competence, etc., and then classify as nouns the words that denote these things: *clock, cloud, cuckoo, colour, chance*, etc. The problem is that the concept of 'thing' implied is far too vague to provide a workable diagnostic.

Bloomfield (1933:266) gives a relevant example: combustion is a process of rapid oxidation producing radiant heat, clearly something that happens rather than a thing or substance, yet words like *fire* and *combustion* are not verbs but nouns. Similar points could be made concerning any number of other nouns: *absence, economy, failure, improvement, lack, probability, similarity, tradition, truth*, and indefinitely many others.

Notice, moreover, that *thing* is the singular form of a count noun, whereas many nouns do not have a count singular interpretation — words like singular noncount *baggage, clothing, cutlery, furniture, lack, machinery, underwear*, or plural noncount nouns like *amends, auspices, regards, remains*, or *spoils*. Nouns like these cannot be said to be names of things: underwear, for instance, is not a thing you wear; amends are not things you make.

Criteria for category membership within a language have to be defined in a very different way, on the basis of appropriate grammatical criteria. For example, the most distinctive property of English nouns is that they function as Head of phrases — NPs — that in turn most typically function as Subject or Object of a clause or Complement of a preposition. Within the

NP they take as Dependents various kinds of determinatives, adjectives, preposition phrases, relative clauses, etc. In addition, a large proportion of them exhibit an inflectional distinction between singular and plural, and between plain and genitive case (*boy*, *boys*, *boy's*, *boys'*, or, with an irregular plural, *woman*, *women*, *woman's*, *women's*).

The fact that grammar books nonetheless repeat the traditional semantically-based nonsense so often, and get away with it, suggests that examples of just a few nouns will suffice to enable readers to grasp the distinction between nouns and verbs on the basis of the tacit knowledge of language they already possess. In other words, rather than identifying nouns by using the traditional definition that they are words that name things, people take the concept of thing to be applicable to the meanings of words that they know to be nouns by virtue of their tacit knowledge of the language they speak.

Pronouns and nouns

The category 'pronoun' is generally treated by traditional grammarians as a distinct 'part of speech' quite separate from noun. This misanalysis, partly based on the semantic intuition that a pronoun does not name anything but merely substitutes for a name, reflects the fact that traditional grammar has a different concept of phrase than modern grammars such as *CGEL*. In the traditional sense a phrase must contain more than one word, but this constraint does not necessarily apply to phrases in the modern sense, where a phrase is a constituent intermediate between word and clause in the constituent structure of sentences. In *The doctor has arrived* the Subject has the form of an NP consisting of a determinative and a noun, whereas in *She has arrived* the Subject NP consists of a noun alone — more specifically a noun of the subclass pronoun rather than common noun.

Traditional grammarians do not generally acknowledge the many disjunctions that are needed in the statement of grammatical rules if pronouns are not recognized as a subtype of noun. For it is not just traditional nouns that can take adjectives in attributive Modifier function, it is either nouns or pronouns (*poor old dad*; *poor old me*); it is not just (NPs headed by) traditional nouns that serve as antecedents for reflexive pronouns, but (NPs headed by) either nouns or pronouns (*Physicists think a lot of themselves*; *They think a lot of themselves*); it is not just traditional nouns (or rather noun-headed NPs) that are found as Complements of prepositions, but NPs headed by either nouns or pronouns (*of London*; *of it*); and so on.

CGEL therefore takes pronouns to be a special subclass of nouns, similar to most proper nouns in hardly ever taking articles and only rather rarely taking attributive Modifiers or relative clauses. Indefinitely many uses of the disjunctive term 'noun or pronoun' are thus avoided.

Auxiliary verbs

CGEL takes auxiliaries (passive or progressive *be*, perfect *have*, supportive *do*, and the modals) to be verbs taking clausal Complements, not minor elements accompanying verbs or mere markers of inflectional features. The idea that auxiliaries are not verbs would have seemed alien to Jespersen, but began to emerge in structuralist work by the 1950s. Charles C.

Fries (1952) called verbs "Class 2 words" but consigned auxiliaries to a "Group B" of minor verb-modifying elements.

Chomsky (1957) took essentially the same approach, introducing *have+en* and *be+en* without category labels under a phrasal umbrella category called 'Aux' which also housed occurrences of 'M' (modal auxiliary). Thousands of other works in the subsequent decades made similar assumptions. But this is misguided. The arguments in favour of taking every auxiliary to be a verb functioning as Head of a clause are compelling (*CGEL*, 1209–1220; for more detailed earlier studies see Huddleston 1976, Pullum and Wilson 1977, and Gazdar et al. 1982).

Specifically, the analysis of auxiliaries in *CGEL* (65–66) makes them verbs that take **simple catenative** Complements: non-finite clausal Complements that are not Objects or predicative or ascriptive or specificational, and can be arbitrarily chained (*may have seemed to want to try to appear to have been...*). There are evident structural differences between *ought to be competent* and *thought to be competent*, but there are no grounds for giving them radically different basic phrase structure configurations: *CGEL* takes both *ought* and *thought* to be verbs taking infinitival Complements.

Treating auxiliaries as verbs heading separate clauses allows for the description of certain facts about negation that posed insuperable difficulties for analyses like that of Chomsky (1957). There are three different negations for a clause like *The senator has always taken bribes from lobbyists*, with quite different truth conditions. The tensed auxiliary verb can be negated alone, yielding *The senator has not always taken bribes from lobbyists*, which asserts the existence of some past time points at which the senator did not take lobbyists' bribes. Alternatively, the nonfinite *take* clause can be negated, yielding *The senator has always not taken bribes from lobbyists*, which says lobbyists always failed to bribe the senator. Finally, both clauses can be simultaneously negated, yielding *The senator has not always not taken bribes from lobbyists*, which asserts that the senator's record of incorruptibility has some blemishes.

By contrast, treating auxiliaries as minor verb-modifying elements within a single clause, instead of as verbs heading their own clauses, makes these facts extremely difficult to account for. The right principle seems to be that every auxiliary is a verb heading its own verb phrase and thus (subject to the very tight limits on comprehensibility of multiple negations) provides a potential locus for negation.

Adjectives, determinatives, and the Modifier function

Virtually all traditional grammars make the gross error, repeated in dictionary after dictionary, of defining adjectives as words that have the (semantic) role of modifying (or "qualifying" or "limiting" or "specifying") a noun. Such a definition takes simply occurring before a noun and affecting its meaning contribution in some way as being sufficient to establish membership in the adjective category. Hence *reluctant* is classified as an adjective because of phrases like *reluctant participation*, but so is *growing* (because of *growing participation*), and *acknowledged* (because of *acknowledged participation*), and *student* (because of phrases like *student participation*), and *Harvard* (because of *Harvard participation*), and *your* (because of

your participation), and *the* (because of *the participation*), and so on. Under this definition adjectives become far and away more numerous than any other category: along with all the genuine adjectives the traditional definition demands that this bloated pseudo-category also embrace every article, quantifier, demonstrative, participle, and noun in the language.

CGEL accordingly rejects such definitions, and recognizes that being able to serve as Modifier of a noun is neither a sufficient condition for adjectivehood nor a necessary one: not all words that modify nouns are adjectives (*Harvard* is a noun), and not all adjectives can function as Modifier of a noun (*asleep* can't). Adjectives are characterized by a cluster of grammatical properties, with core members of the category having them all, while more peripheral members have some but not all. Core adjectives share these three properties:

 (i) They can function as Modifier of a noun (*big ideas*).
 (ii) They can function as Predicative Complement (*This is big*).
 (iii) They can be modified by adverbs of degree (*very big*).

The great majority of the core adjectives — an open class including *abstemious*, *brilliant*, *combustible*, *democratic*, *efflorescent*, *fantastic*, *geological*, *hopeless*, *intelligent*, *judicious*, and so on — are modified for degree or comparative grade by adverbs like *very*, *extremely*, *more*, and *most*. A much smaller set including *able*, *big*, *cute*, *dumb*, *easy*, *fine*, *great*, etc., inflect for comparison (*great*, *greater*, *greatest*). Some more peripheral adjectives lack one or more of the properties; for example, *mere* has only property (i), and *asleep* has only property (ii).

CGEL also posits a distinct category of **determinatives**, to which the articles, demonstratives, and quantifiers belong: *a ~ an*, *all*, *every*, *few*, *many*, *most*, *some*, *that*, *the*, *this*, and perhaps a couple of dozen more items. Our name for this category follows Palmer and Blandford (1939), who relate it to the French grammatical term 'adjectif determinatif'; see also Huddleston (1984). But there is some unfortunate variation in the terminology found in this area. *CGEL* uses 'determinative' as a category term (with the suffix *-ive* matching that in the category term 'adjective') and 'Determiner' as a function term (with the suffix *-er* matching that in the function term 'Modifier'). Quirk et al. (1985) do the opposite: for them 'determinative' is a function term, and 'determiner' a category term. Many works don't use 'determinative' at all, and take 'determiner' to be a category term.

Again, as with adjective and Modifier, it is crucial not to confuse the category with the function. Serving in (what *CGEL* calls) Determiner function is neither a sufficient nor a necessary condition for belonging to the determinative category. It is not sufficient because it is not just determinatives that can serve as Determiner of an NP (as in *all responsibility*), but also NPs in the genitive case (as in *the tenant's responsibility*). And it is not a necessary condition because determinatives can often serve in Modifier function: while the NP *one problem* has the determinative *one* in Determiner function, *the one problem we face* has it as Modifier of the noun *problem*; and while *the reason* has *the* as Determiner, a phrase like *all the better for it* has *the* functioning as Modifier of a comparative adjective.

The forms *my*, *your*, *his*, *her*, *its*, *our*, *their*, and *one's* are traditionally treated as 'possessive adjectives', but this is another error by earlier grammarians. They are neither

adjectives nor determinatives: they can serve as Determiner simply because they are the dependent genitive forms of the pronouns *I*, *you*, *he*, *she*, *it*, *we*, *they*, and *one*, and NPs headed by dependent genitive forms of nouns always function as Determiner. (Independent genitives like *mine*, *yours*, *his*, *hers*, *its*, *ours*, and *theirs* are pronouns that function as Head of a plain-case NP, which can never function as Determiner: *mine idea.)

CGEL avoids the term 'possessive', even though genitives sometimes express a possession relation of one kind or another. Notice that *the tenant's responsibility* does not involve any reference to possession or ownership by the tenant; the genitive case has a set of uses that go far beyond possession in any ordinary sense. There is surely no literal owning or possessing implied in <u>*Mike's*</u> *sister*; <u>*the dollar's*</u> *strength against the euro*; <u>*this plan's*</u> *chief failing*; <u>*his having fallen ill*</u>; <u>*high speed rail's*</u> *huge initial cost*; <u>*the prisoner's*</u> *disgraceful treatment at the hands of the police*; etc.

Prepositions, adverbs, and subordinators

Prepositions, as we mentioned above, are defined in all traditional grammars as words that relate one noun to another. Yet quite uncontroversial prepositions like *as*, *at*, *from*, *of*, *on*, *until*, and many others can clearly be followed not only by NPs but also in some cases PPs, adjectives (or adjective phrases), adverbs (or adverb phrases), gerund-participial clauses, infinitival interrogative clauses, finite interrogative clauses, or finite exclamative clauses:

It came from <u>under that couch</u>.	(preposition phrase)
I see it as <u>quite obvious</u>.	(adjective phrase)
They did not realize until <u>somewhat later</u>.	(adverb phrase)
He's been depressed since <u>leaving the city</u>.	(gerund-participial clause)
They couldn't agree on <u>what to buy</u>.	(infinitival interrogative clause)
It depends on <u>how much you can pay</u>.	(interrogative content clause)
I'm aware of <u>what a shock this will be</u>.	(exclamative clause)

It is also clear that uncontroversial prepositions like *in*, *up*, *down*, *over*, *through*, etc., are sometimes not followed by a noun (or NP) or anything at all:

Soon they went <u>in the house</u>.	*Soon they went <u>in</u>.*
He came running <u>up the street</u>.	*He came running <u>up</u>.*
Does this hole go right <u>through the wall</u>?	*Does this hole go right <u>through</u>?*
They like to run <u>around the yard</u>.	*They like to run <u>around</u>.*

To take account of such facts, traditional grammar posits that a substantial subset of the prepositions have homophonous and virtually synonymous doppelgangers belonging to other categories. Thus they recognize the word *down* as a preposition in *He fell down the steps* (it is followed by the NP *the stairs*) but not in *He fell down on the steps*. In the latter, since there is no NP following it, *down* has to be an adverb (it does, after all, modify the verb *fell*). Another (overlapping) subset are alleged to have doppelgangers in the 'subordinating conjunction' category: in *before her court appearance* the word *before* is acknowledged as a preposition, but in *before she appeared in court*, where what follows *before* is a declarative content clause, it is claimed not to be.

There is no semantic, morphological, or phonological support for having three words spelled *before*. Yet traditional grammars (and all published dictionaries we know of) say *before* is a preposition in *I never saw her before her court appearance*, a 'subordinating conjunction' in *I never saw her before she appeared in court*, and an adverb in *I never saw her before*). In each case *before* heads a phrase functioning as a temporal Modifier; the pronunciation is the same, the morphology is the same (it is uninflectable), and the meaning is the same (considering just the temporal sense, it refers to the period preceding a designated point in time identified either by the interpretation of the Complement — the event time if it denotes an event, or the reference time if it denotes a proposition — or by the context when the preposition phrase consists of simply the preposition as Head).

CGEL is the first reference grammar to take note of cogent but long neglected arguments in such works as Hunter (1784), Jespersen (1924), Geis (1970), Emonds (1972), and Jackendoff (1973, 1977) in favour of unifying the category assignments for these prepositions. Words like *after*, *before*, and *since* are treated as prepositions not only when they have an NP as Complement but also when the Complement is a clause or when there is no Complement.

This amounts to extending the selection of **subcategorization** of lexical heads (i.e. classifying them according to what categories can function as their Complement) from the case of verbs (where it is traditional, and everyone agrees it is appropriate) not just to nouns and adjectives (where few grammarians dispute it, though complementation of nouns and adjectives is little noted in traditional grammars) but to prepositions as well.

In fact *CGEL* also extends complementation to genuine adverbs, refuting the proposal in Jackendoff (1977:78) that the key difference between adjectives and adverbs is that adjectives take Complements and adverbs do not. *CGEL* cites several small classes of adjective-derived adverbs that take PP Complements in the same way that their related adjectives do:

independently of these considerations	(*of*-PP Complement)
fortunately for the others	(*for*-PP Complement)
differently than anyone else had	(*than*-PP Complement)
similarly to its counterparts in other industries	(*to*-PP Complement)
separately from the rest of the company	(*from*-PP Complement
simultaneously with the rebellion in the south	(*with*-PP Complement)

Subordinate clause types

The traditional analysis of finite subordinate clauses distinguishes three subcategories which are supposed to be syntactically analogous to the lexical categories noun, adjective, and adverb. 'Noun clauses' are supposed to be the ones that can function as Subject or Object just as NPs can; 'adjective clauses' are those that modify nouns the way adjectives supposedly do; and 'adverb clauses' are clauses that, like adverbs, semantically modify verbs.

Such an analysis fails comprehensively. It is based on analogies that are essentially all spurious. Since this is a major departure, we will summarize the reasoning here (a fuller presentation is given in Huddleston and Pullum 2004).

To begin with, note that the putative subcategories of subordinate clause overlap: indefinitely many clauses have to be treated as belonging to two distinct subcategories. In *I was told that I kissed her* the subordinate clause would be classified as a 'noun clause' on grounds that it is analogous to the direct Object of *tell*, as in *I was told a lie*. But in *I was so bold that I kissed her* the identical clause would be classified as an 'adverb clause' of result.

What is wrong here is that the classification is trying to attribute to the clauses themselves what is really a difference in the functions that a subordinate clause can serve: the difference between Complement and Adjunct. Classifying *that I kissed her* as belonging to two different subcategories of clause is no more sensible than treating *that day* as belonging to two different subcategories of NP in *I spent that day alone* (where it is an Object) and *I was alone that day* (where it is a temporal Adjunct).

The traditional analysis of subordinate clauses is intimately linked to the miscategorization of those prepositions that take clauses as Complement. As noted in the previous section, prepositions clearly take gerund-participial clauses (*since leaving the city*), infinitival interrogatives (*agree on what to buy*), finite interrogatives (*depends on how much you can pay*), and finite exclamatives (*aware of what a shock this will be*). Yet on encountering a preposition followed by a finite declarative clause (as in *since he left the city*), the traditional account abandons the obvious analysis and claims that *since* belongs with the 'subordinating conjunctions' like *that* or *whether*. The combination of the preposition with the clause is then referred to as an 'adverb clause'.

The traditional account faces the embarrassing fact that in exactly those cases where a PP with *since* can be found (e.g. *Have you spoken to him since his resignation?*) the alleged 'adverb clause' can also appear (*Have you spoken to him since he resigned?*). The analysis provides no reason to think this would happen: phrases headed by the noun *crease* and the verb *crease* do not have anything like the same distribution. Why would a PP headed by the preposition *since* be found in just the same places as an 'adverb clause' introduced by a 'subordinating conjunction' with which it has no grammatical connection?

The traditional account is a gross mistake. It is irrational to single out finite declarative clauses and keep them off the list of categories that can be the Complement of a preposition. And once we see that *since he resigned* is best treated as a PP, it is easy to see that the same sort of account is also best for *although she didn't mean it*, *after the thrill is gone*, *because you come to me*, *before I met you*, *though nobody knew*, *unless you come back*, etc. They should all be recognized as preposition phrases (PPs).

Most prepositions that take finite clause Complements take bare finite clauses like *he resigned*, not clauses expanded by the default subordinating marker *that*. But there are exceptions: prepositions like *except*, *given*, *notwithstanding*, and *provided* take expanded clauses that do have the subordinating marker (*Given that you obviously don't like us, I'm surprised you're here*).

However, *CGEL* does not treat the clause-subordinating unstressed *that* as a preposition. Notice that it is often optional: *He believes they despise him* is just as grammatical as *He believes that they despise him*, and has the same meaning. This is quite unlike the behaviour of prepositions: the Head preposition in a PP like *since they despise him* is never optional. *That* is a meaningless default subordinating marker for the default type of content clause (declarative), which makes it less surprising that it is omissible in the most typical context for a subordinate clause, immediately following a verb (see *CGEL* 952–4). And its interrogative counterpart *whether*, though not omissible, is in many contexts replaceable by a slightly more informal alternant, interrogative *if*, as in *I wonder if it's true*.

CGEL categorizes *that*, *whether*, and interrogative *if* as **subordinators**, but claims that all the other traditional 'subordinating conjunctions' are prepositions. (See Emonds 1985 for an attempt within transformational grammar to push in the direction of fully unifying subordinators with prepositions.)

Turning now to the three analogies with lexical categories that lie at the heart of the traditional classification of finite subordinate clauses, we find that all three collapse rapidly on close examination.

It is very clear that noun clauses do not function like nouns. With verbs one might be tempted to see an analogy between clauses as Complement and NPs as Object (e.g. to regard *I regret that I lied* as parallel to *I regret my lie*), but clauses are also found as Complement with adjectives, hardly any of which can take NPs (*I'm sorry that I lied* vs. **I'm sorry my lie*), and as Complements of nouns, which absolutely never take NP Complements (contrast *my regret that I lied* with **my regret my lie*).

There is a traditional way of sidestepping this problem: saying that in *my regret that I lied* the clause is "in apposition" to the noun *regret*. But there are several reasons why this would not be a viable solution. First, it wrecks the parallelism with *I regret that I lied*, where there is no hint of an appositional relation with anything. Second, it fails to explain why we can't say **my regret my lie*, with the NP *my lie* the apposed element (compare *her brother the heart surgeon* where the element in apposition is an NP). Third, it fails to explain why **my dishonesty that I lied* isn't grammatical (it should be just another case of a clause in apposition to a noun). Fourth, it fails to explain why the clause in *my regret that I lied* cannot stand in place of the whole construction (in *My boy Jack will see to it* the appositional NP *Jack* can replace the containing constituent, yielding *Jack will see to it*, but in *My regret that I lied was genuine* the analogous substitution yields ungrammaticality: **That I lied was genuine*).

A different sidestepping move, to cover cases of 'noun clause' Complements of adjectives that never take NP Complements, is to posit 'suppressed' prepositions; that is, to treat *We're pleased that you got here* as parallel to *We're pleased at/over your arrival*, and so on. But for cases like *I'm afraid you're wrong* there is no possible PP to analogize from (**I'm afraid {of / at / over / with / from} your error*). As Goold Brown (1851:597) astutely remarked concerning the hypothesizing of an underlying element that is not allowed to appear overtly, "where it cannot be inserted without *impropriety*, it is absurd to say, that it is 'understood'."

The point is that different nouns, verbs, and adjectives licence different categories as Complement, and they cannot be reduced to each other. Each of the eight ways in which a

lexical item could logically be specified for taking NP and/or PP and/or content clause Complements is instantiated in English for at least some items.

Further evidence against the traditional notion of 'noun clause' comes from extraposition: *That nobody knows how to repair them is unfortunate* has the stylistic alternant *It is unfortunate that nobody knows how to repair them*, but *The widespread ignorance about how to repair them is unfortunate* does not likewise have the alternant **It is unfortunate the widespread ignorance about how to repair them*.

In short, attempting to derive the distributional properties of finite subordinate clauses from claims about their being functional equivalents of nouns, adjectives, or adverbs cannot work. The traditional account fails in just about every way possible.

CGEL classifies subordinate clauses on the basis of their internal structures instead. The basic division is between **content clauses**, **relative clauses**, and **comparative clauses**. Content clauses are much like independent clauses, with the exception that (i) they are never imperative (imperative clauses are always main clauses), (ii) they may have subordinators introducing them, and (iii) the open interrogative type does not exhibit auxiliary-before-Subject structure in Standard English (we find *I wonder who it is* rather than **I wonder who is it*).

Relative clauses are similar except that they obligatorily contain a 'gap' — an unfilled NP or PP position inside them — and they may be introduced by a relative phrase like *which* or *who*. Thus among the relative clauses comparable to the main clause *He was reading a book* (with the gap shown by '∅') are the underlined parts of *the book <u>which he was reading ∅</u>, the book <u>that he was reading ∅</u>*, and *the book <u>he was reading ∅</u>*.

Comparative clauses, appearing mainly as Complements to *as* or *than*, also have reduced internal structure: in *You're nearly as stupid as they are ∅* the predicative constituent normally needed to follow *are* is missing, and in *I've read more books than you've had ∅ hot dinners* the quantificational Determiner of *hot dinners* is obligatorily absent (notice that inserting one causes ungrammaticality: **I've read more books than you've had many hot dinners*).

Relative clauses and comparative clauses have to be distinguished both from content clauses and from each other. The distinguishing properties are real, they involve differences in internal syntactic structure, those differences have correlates in external distribution, and syntactic generalizations are captured if we classify subordinate clauses in this way. None of this can be said about the traditional classification.

Discourse and information presentation

For many linguists the main interest of the study of syntax lies in the structure of the constructions that show sensitivity to discourse context: preposing, postposing, clefting, existentials, passives, and so on. It is remarkable how little interest traditional grammarians showed in such phenomena: typical pedagogical grammars barely mention them. *CGEL* describes them in detail (in Chapter 16) and stresses that the role they play lies in the

packaging and presentation of the information contained in a sentence, and that they have very clear syntactic properties that do not arise simply from common sense or natural communicative inclinations.

Crucially, the differing word orders that these constructions allow cannot be treated as a matter of utterer's discretion: in Saussure's terminology, they cannot be treated as **parole** rather than **langue**. There are intricate and quite specific syntactic constraints to be described, many of which were not appreciated until the work of transformational-generative grammarians brought them to light in the 1960s and 1970s.

The preposing of phrases in Complement function provides a simple example. The word order difference between *I certainly liked your latest article* and *Your latest article, I certainly liked* might appear at first glance to be nothing more than a matter of whim: the utterer chose to produce the Object of the clause before starting on the Subject NP and the verb. That is what we mean by the parole view. And it is clearly wrong.

First, it is not just Objects within a clause that this preposing can apply to; all sorts of subconstituents can be preposed, leaving a gap in place of the Complement in a PP, or the Object in a subordinate clause, or the Complement within a preposition in a subordinate clause, or any of an arbitrary number of other possibilities (again, we mark the gap with '∅', and show content clause boundaries with square brackets):

Some of his remarks I approved of ∅.	(Complement in PP)
Some of his remarks I imagine [the audience enjoyed ∅].	(Object in content clause)
Some of his remarks I imagine [the audience approved of ∅].	(Complement in PP contained within content clause)
Some of his remarks I imagine [his colleagues might decide [they think [he should retract ∅]]].	(Object inside content clause inside content clause inside content clause)

The construction thus involves what theoretical linguists often call an **unbounded dependency**, as found also in open interrogatives (*What do you imagine [his colleagues might decide [they think [he should retract ∅]]]?*) and relative clauses (*the claim that you imagine [his colleagues might decide [they think [he should retract ∅]]]*).

Second, there are limits to where the gap may be: although the preposed Complement can be the Subject of an embedded content clause, it cannot be a Subject that immediately follows a subordinator, or a Coordinate within a coordination, or a constituent contained within one of the Coordinates of a coordination:

> Some of his remarks
>
> I imagine [∅ surprised them]. (Subject of content clause)
>
> *Some of his remarks I imagine (Subject immediately following
> [that ∅ surprised them]. subordinator: not grammatical)
>
> *Some of his remarks I imagine (Coordinate of the underlined NP
> [they enjoyed ∅ and the music]. coordination: not grammatical)
>
> *Some of his remarks I imagine [they ignored (Object in Coordinate of the underlined
> the rest of the lecture but enjoyed ∅]. VP coordination: not grammatical)

These are syntactic constraints on the entire range of unbounded dependency constructions. Figuring out exactly how those constraints need to be stated is a subtle and complex matter that may in part be psycholinguistic rather than purely linguistic (the work of Hofmeister and Sag 2010, for example, shows that certain processing difficulties have sometimes been mistaken for syntactic constraints).

Third, it is clear that a clause with Complement preposing cannot be deployed in a discourse in the same ways as its more basic counterpart. A sentence such as *Your latest article I certainly liked* would be bizarre as the opening sentence of a conversation, for example. Complement preposing is appropriate where it picks up on a topic just introduced or highlights some kind of a focused contrast with some similarly preposed Complement. In this sentence we see both factors at work:

> *I'm afraid I wasn't able to be complimentary about all of your work. Your book I liked a lot, but your most recent article I regarded as unconvincing.*

In similar ways, *CGEL* discusses a variety of other constructions in the context of their discourse-sensitive information-packaging roles: postposing (*Turning around, he saw on the desk a gun*) and inversions of subject and dependent (*There goes the neighbourhood*); existential and presentational clauses (*There was a full discussion*); extraposition (*It amazes me that you haven't left him*); dislocations (*He's quite a handful, our Jimmy*), clefts (*It was only later that he figured it out*), and the various constructions known as passives, together with the pragmatic conditions favouring or disfavouring their use.

Technicalities

The standard representation of sentence structure in modern linguistics is the ordered, node-labelled tree. Strictly speaking, *CGEL* departs from this standard representational formalism, in two distinct ways.

First, the inclusion of function labels as well as category labels is best understood formally in terms of graphs with labels on the edges as well as the vertices. That is, while the nodes bear category labels as with ordinary trees, the lines are also labelled, from an inventory of function labels. Purely for typographical convenience, the diagrams in *CGEL* have the function labels moved down to the node below and prefixed to the category labels with a

colon as separator. Thus a diagram depicting a clause node immediately dominating a node labelled 'Subject:NP' (intuitively representing an NP functioning as the Subject of a clause) should be taken to be merely a different visual presentation of a graph in which the downward line from a clause node to an NP node is labelled 'Subject'.

CGEL's second departure from using trees involves a tightly restricted class of cases in which two edges are allowed to converge on a single node in a downward direction from the root. Under the standard definition of a tree, a node has either no parent (in the case of the root node) or exactly one parent (all other nodes). Three classes of phenomena are treated in *CGEL* using this device:

(i) Most determinatives (e.g. *all*, *each*, *few*, *many*, *most*, *none*, *several*, *some*, *that*, *this*, etc.) can stand alone as NPs, as in *All are included here*; *None were saved*; *Several decided not to bother*. *CGEL* posits that in such cases the determinatives bear two functions, serving as Determiner and Head simultaneously.

(ii) Many adjectives can appear in NPs with the definite article but no head noun, being interpreted as if they had a head noun with either human (plural) or abstract generic sense: *The good die young*; *The poor are always with us*; *He respected even the humblest of them*; *I blame the French for this*; *It's a leap into the unknown*. *CGEL* treats these as cases of an adjective taking on the function of Head in addition to the function of Modifier.

(iii) In relative constructions such as the underlined NP in *What Frankenstein created would later destroy him* we treat the initial *what* as both the head of the NP and the preposed *wh*-phrase in the relative clause.

This downward convergence of edges is called **function fusion** in Payne, Huddleston and Pullum (2007), which provides an extended defence and a new application of the device.

Function fusion constitutes a second way in which *CGEL* does not strictly employ trees as its structural representations throughout. However, the departure from trees is actually very slight, in a sense that can be made formally precise. The function labels on lines could in principle be eliminated by simply cross-multiplying the category and function inventories (e.g., treating 'Subject:NP' for mathematical purposes as an unanalysable unit). And the departure from treehood implicit in function fusion is purely local: in *CGEL*'s analyses, whenever two constituents *A* and *B* share a single immediate constituent *C* in some fused function, either *A* = *B* (as when ordinal *second* is both Modifier and Head of the nominal node in *the second*: *CGEL* p. 412, [7c]) or *A* and *B* are the grandparent and the parent of *C* (as in the NP *someone I know*, where *someone* is both Determiner and Head: *CGEL* p. 412, [7b]).

Given these two facts together with certain other plausible conditions that seem very likely always to be met, Pullum and Rogers (2008) showed that the expressive power of a grammar placing conditions on *CGEL*'s partially tree-like graphs will be the same as it would be if trees had been employed throughout. That is, languages that are describable using *CGEL*-style non-tree graphs could also be given an alternative description using trees if this were desired.

The use of node-sharing representations that is made in *CGEL* can thus be seen as a heuristic decision: it represents certain facts perspicuously, capturing the right generalizations

about them, but if necessary for some theoretical or computational purpose a strongly equivalent description could be given entirely in terms of constraints on trees.

Conclusion

CGEL follows neither traditional grammar nor modern theoretical linguistics in its approach to the description of English. Its conception was in part inspired by previous large-scale efforts like those of Jespersen (1909-1949) and Quirk et al. (1985), and in part informed by a large body of syntactic discoveries during the post-1957 era in linguistics, but it does not employ the theoretical apparatus of any of these.

What *CGEL* is attempting to do is to present a synthesis of the clearest ideas about describing English grammar that have been developed over the past 400 years, and to provide detailed coverage of the full array of facts that any adequate grammar of English would have to deal with.

This chapter has stressed in particular the departures from traditional presentations: recognizing the status of phrases in the modern sense; folding pronouns in as a special subcategory of nouns; treating auxiliaries as a special subcategory of verbs; separating the category of adjectives from both the small category of determinatives and the broader class of items that can function as Modifier; delimiting prepositions in a way that recognizes their different patterns of complementation but also separates them from the very small category of subordinators; revising the analysis of subordinate clause types; and making explicit the ways in which a number of syntactic constructions are sensitive to discourse and information structure.

The chapter has also briefly touched on the technical details of the implicit syntactic theory behind *CGEL*, which we think is best visualized as involving conditions imposed on richly labelled structure-representing graphs.

References

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Bullokar, William. 1586. *Pamphlet for Grammar*. London: Henry Denham.

Brown, Goold. 1851. *The Grammar of English Grammars*. New York: Samuel S. and William Wood.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Dryden, John. 1672. In defense of the epilogue; or, an essay of the dramatique poetry of the last age. Appended to *The Conquest of Granada by the Spaniards*. London: Henry Herringman.

Emonds, Joseph E. 1972. Evidence that indirect object movement is a structure-preserving rule. *Foundations of Language* 8: 546–561.

Emonds, Joseph E. 1985. *A Unified Theory of Syntactic Categories*. Dordrecht: Foris.

Fries, Charles C. 1952. *The Structure of English*. New York: Harcourt Brace.

Gazdar, Gerald, Geoffrey K. Pullum, and Ivan A. Sag. 1982. Auxiliaries and related phenomena in a restrictive theory of grammar. *Language* 58: 591–638.

Geis, Michael. 1970. *Adverbial Subordinate Clauses in English*. Ph.D. thesis, MIT.

Gwynne, N. M. 2011. *Gwynne's Grammar: The Ultimate Introduction to Grammar and the Writing of Good English*. London: Idler Books.

Harvey, Thomas W. 1868. *A Practical Grammar of the English Language for the Use of Schools of Every Grade*. New York and Cincinnati: Van Antwerp, Bragg & Co. (Revised edition 1878.)

Hofmeister, Philip and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Huddleston, Rodney D. 1976. Some theoretical issues in the description of the English verb. *Lingua* 40: 331–383.

Huddleston, Rodney D. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.

Huddleston, Rodney D. 1988. Review of Quirk et al. (1985). *Language* 64: 345–354.

Jackendoff, Ray S. 1973. The base rules for prepositional phrases. In Stephen R. Anderson and Paul Kiparsky, eds., *A Festschrift for Morris Halle*, 345–356. New York: Holt Rinehart and Winston.

Jackendoff, Ray S. 1977. *X′ Syntax*. Cambridge, MA: MIT Press. [NOTE: *X′* is to be typeset as capital *X* with a macron or overbar; can't do that in Microsoft Word!]

Jespersen, Otto. 1909–1949. *A Modern English Grammar on Historical Principles*. London: George Allen and Unwin.

Jespersen, Otto. 1924. *The Philosophy of Grammar*. New York: Holt.

Klima, Edward S. 1965. *Studies in Diachronic Syntax*. Ph.D. thesis, MIT, Cambridge, MA.

Linn, Andrew. 2006. English grammar writing. Chapter 4 of *The Handbook of English Linguistics*, ed. by Bas Aarts and April McMahon, 72–92. Oxford: Basil Blackwell.

Murray, Lindley. 1795. *English Grammar Adapted to the Different Classes of Learners*. York: Thomas Wilson and Sons.

Nesfield, J. C. 1900. *Outline of English Grammar*. London: Macmillan.

Palmer, Harold E. and F. G. Blandford. 1939. *A Grammar of Spoken English on a Strictly Phonetic Basis*. Cambridge: W. Heffer & Sons. 2nd edition of a book first published in 1924.

Payne, John; Rodney Huddleston; and Geoffrey K. Pullum. 2007. Fusion of functions: The syntax of *once*, *twice*, and *thrice*. *Journal of Linguistics* 43: 565--603.

Pullum, Geoffrey K. 1974. Lowth's grammar: a re-evaluation. *Linguistics* 137: 63–78.

Pullum, Geoffrey and Deirdre Wilson. 1977. Autonomous syntax and the analysis of auxiliaries. *Language* 53: 741–788.

Pullum, Geoffrey K. and James Rogers. 2008. Expressive power of the syntactic theory implicit in *The Cambridge Grammar*. Presented to the Linguistics Association of Great Britain (online at http://www.lel.ed.ac.uk/~gpullum/EssexLAGB.pdf).

Quirk, Randolph; Sidney Greenbaum; Geoffrey Leech; and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Tieken-Boon van Ostade, Ingrid. 2011. *The Bishop's Grammar: Robert Lowth and the Rise of Prescriptivism in English*. Oxford: Oxford University Press.

Wallis, John. 1653. *Grammatica Linguae Anglicanae*. Hamburg: Gotfried Schultzen.