Notes on the Cambridge Grammar of the English Language

Gihan Marasingha

August 31, 2023

1 About this document

These are my notes on *The Cambridge Grammar of the English Language* (CGEL).

2 Grammar as a description of English

A grammar of a language is one that aims to 'describe the principles or rules governing the form and meaning of words, phrases, clauses and sentences'. It has two parts **syntax** and **morphology**. The former being concerned with the way in which words form larger units (phrases, clauses, sentences) and the latter with the construction of words themselves.

CGEL also touches on some topics outside syntax and morphology. It examines **semantics**, distinguishing between **grammatical semantics**, this being meaning that follows simply from the language system) and **pragmatics**, this being about meaning in particular situations.

Unlike mathematical definitions, those in CGEL are not stipulative. Rather, they are descriptive. They must be, for the grammar contains no undefined terms.

Thus, where CGEL defines the grammatical category *noun*, it does so by describing a set of characteristics of **prototypical members** (also called the **central** or **core** members) of the category. This is to acknowledge that some words that we class as nouns do not satisfy all the attributes specified in the grammar.

For example, *grape* is a prototypical noun: it denotes a physical object, it has an inflectional form contrast between singular and plural, it functions as the head of a noun phrase. However, the noun *equipment* shows no such inflectional form.

Likewise, in *race discrimination*, the noun race does not act as head of a noun phrase. Instead, it is a modifier of the noun *discrimination*. However, this does not make *race* an adjective. The word race does not satisfy many of the other characteristics of adjective. For example, it cannot be used predicatively: *I am race* is ungrammatical.

Put another way, the grammar is not sufficient, by itself, to determine what is and what is not a noun. This discrimination can only be established by something external: the observation of how we use particular words or phrases in practice.

My thought on the matter is that a grammar is simply descriptive in the same sense that Linnaeus' taxonomy is descriptive. In each case, there is something *out there* that we are attempting to codify. But the thing itself (be it the relationships between biological species or the construction of English language sentences) has no intrinsic *formal* structure. There will always be 'edge cases' that are difficult to pin down.

3 The components of the grammar

As mentioned, the main components of the grammar are the study of syntax and the study of morphology. We briefly treat each of these in turn.

3.1 Syntax

The theoretical framework for syntax in CGEL is founded on three principles:

1. Sentences have parts, which may themselves be parts.

2. The parts of sentences belong to a limited ranges of types.

3. The parts have specific roles or functions within the larger parts they belong to.

These three principles correspond and lead to the notions of **constituent structure** analysis, **syntactic categories** and **grammatical functions**.

3.1.1 Constituent structure

Consider the foundational notion that a sentence can be divided into parts, each of which may also be divided into parts, called **constituents**.

Figure 1 shows one way to decompose the clause, *a bird hit the car* as a hierarchy of constituents.
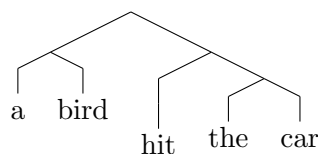

Figure 1: A syntax tree for a simple sentence

In this diagram, the phrases *a bit* and *hit the car* are the **immediate constitutions** and the leaf nodes (here words) are the **ultimate constituents** of the sentence.

Evidently, this is not the only possible way to impose a hierarchy on (equally to *analyse*) *a bird hit the car*. That this is the *correct* analysis can only be appreciated by bringing to bear other aspects of the grammar.

3.1.2 Syntactic categories

We have recognised that sentences can be recursively analysed into 'parts'. We have claimed that each part belongs to a limited ranges of categories. In this section, we briefly examine those categories.

First, we divide the class of categories in two: the **lexical categories** and the **phrasal categories**. Every ultimate constituent (i.e. word) of the sentence belongs to a lexical category. A higher-level structure that depends on a central word and which may contain more than one word is a phrasal category.

Lexical categories The lexical categories consist of the 'parts of speech'. Principal among these are: noun (N), verb (V), adjective (Adj), adverb (Adv), preposition (Prep), determinative (D), subordinator, coordinator, and interjection.

Thus, *a* is a determiner and *bird* is a noun. The syntax tree in Figure 2, gives lexical-category labels for each of the words in *a bird hit the car*.
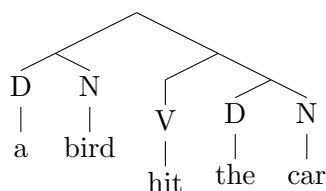

Figure 2: A syntax tree with lexical category labels

Phrasal categories A **phrase** is a constituent at a higher level than the lexical level. Each phrase typically contains a 'most important word' that determines its type. It may also contain other constituents that serve to elaborate its meaning.

For example, a phrase consisting of a noun and supporting constituents is a **nominal**; a nominal together with a determinative is a **noun phrase**; a verb with its supporting constituents is a **verb phrase**; a noun phrase with a verb phrase is a **clause**.

The principal phrasal categories are: Clause (Clause), verb phrase (VP), noun phrase (NP), nominal (N), adjective phrase (AdjP), adverb phrase (AdvP), preposition phrase (PP), and determinative phrase (DP).

Figure 3 supplements our previous diagram with the inclusion of phrasal category labels.

3.2 Grammatical function

The last of our three foundational pillars of syntax is **grammatical function**. The function of a constituent denotes its relationship either to the larger construction containing it to to another element within that construction. The term 'construction' is not defined in CGEL. I take it to mean a phrase when considered as a structure whose elements are related to each other and which may be related to higher-level structures.
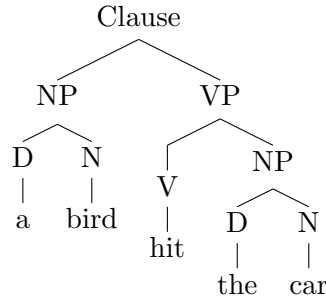
Clause

Figure 3: A syntax tree with phrasal and lexical category labels

The major two types of function are *head* and *dependent*. In the noun phrase *a bird*, the noun *bird* is the head and the determinative *a* is a dependent of the construction.

Each construction has at most one (and usually exactly one) **head**, its most important element. Generally, the category of a phrase depends on that of the head: a phrase with a noun as a head is a noun phrase, a phrase with a verb phrase as a head is a clause.

A phrase can contain more than one dependent. Consider the verb phrase *gave the key to the landlord*. Its head is the verb *gave*. It has two dependents: *the key*, a noun phrase and *to the landlord*, a preposition phrase.

Depending on the category of the construction at hand, we assign special names to the head and dependents. For instance, a canonical clause has two constituents: a **subject** and a **predicate**. Here, predicate is just a special case of head and subject is a special case of dependent. The subject canonically takes the form of a noun phrase and the predicate a verb phrase.

We see this in Figure 4 where *a bird* is the subject (dependent) and *hit the car* is the predicate (head) of the clause. We use a triangle (or roof) over a constituent to show that the analysis is not complete to the lexical level.

Diagramatically, it would make most sense to display grammatical function labels on the edges of the syntax tree, but for the sake of convenience, we instead label the nodes.
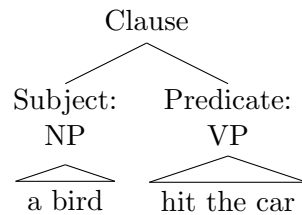
Clause

Figure 4: A syntax tree with functional and phrasal labels

A phrase can have more than one dependent. In Figure 5, we have a syntax tree for the verb phrase *gave the key to the landlord*. The head of a verb phrase is called a predicator and is a verb. Here, the verb *gave* is the head (predicator). Here the verb phrase has two dependents: one an 'object' and one a 'complement'.
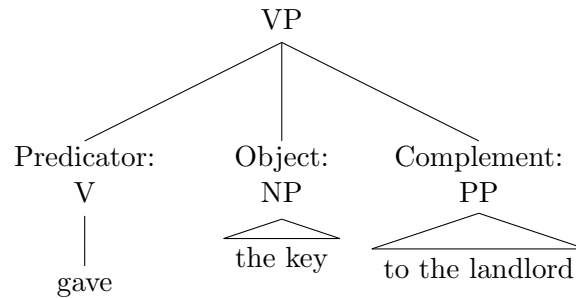
4

Figure 5: A phrase with two dependents

The previous example shows an important distinction between syntactic categories and grammatical functions. Whereas the phrase *the key* is always a noun phrase in any context, its grammatical function depends fundamental on the context. In the clause, *the key is shiny*, the phrase *the key* is a subject. But the same phrase is an object in the clause, *Bob stole the key.*

Not every construction has a head. The phrase *fish and chips* is, syntactically, a NP-coordination. This is syntactic structure we have not yet seen so far. It is a NP and also a **coordination**, this being a relation between two or more elements of syntactically equal status, the **coordinates**, usually linked by a **coordinator** such as *and* or *or*.

Notably, *fish and chips* has no head. Instead it has two dependents, the coordinate *fish* and the coordinate *and chips*, each of which is a NP. The phrase *and chips* can be further analysed into the marker *and* (a coordinator) and the coordinate *chips* (a NP). An example similar to this, *Kim and Pat* is diagrammed in Chapter 11, Section 1.1 of CGEL.

3.3 Morphology

Morphology is concerned with the shape of words and lexemes. Exactly what constitutes a word isn't made explicit in CGEL, except to say that a **word** is the smallest unit of syntax. As I understand it, this means that two tokens are instances of the same word if they are identical as strings and are syntactically equal.

Thus, the occurrences of *dog* in *I fed the dog* and *the dog fed me* represent the same word. Though *dog* is (head of) the object in the first phrase and (head of) subject in the second and thus the occurrences show a different in *grammatical function*, from a purely *syntactic* point of view, they are identical. They are both nouns.

It is still unclear to me whether the two occurrences of the string *read* in the sentences *I read the book* and *Will you read the book?* represent the same word. They are identical as strings, they both belong to the lexical category *verb*, but one is a present tense and the other is a past tense. I don't believe CGEL refers to tense as a syntactic category, but it is referred to as a syntactic property.

More simply, *walk* and *walked* denote different words as they are different strings, even though they are clearly related. We say instead that *walk* and *walked* are forms of the

same **lexeme**.

Here, a lexeme is a set of words that represent roughly the same notion, but exhibit **inflection** (a change of form) to reflect some syntactic property. Here, *walk* is a present tense **inflectional form** and *walked* is a past tense inflectional form of the same lexeme.

In CGEL, a bold emphasised font is used to denote the lexeme represented by the given word. Thus, **walk** is the lexeme of which *walk* is an inflectional form. ' The study of morphology has two aspects: **inflectional morphology** and **lexical word-formation**.

Inflectional morphology is concerned with how a **lexical base** is modified to give the inflectional form corresponding to a given syntactic property. Thus the lexical base *car* is modified with the suffix ·*s* to give the inflectional form *cars*. This reflects the value of the syntactic property of *number* which here is *plural*.

In contrast, lexical word-formation is concerned with the construction of lexical bases. This can happen in one of three ways. The lexical base *bluebird* is formed by **compounding** the bases *blue* and *bird*; the base *quickly* is a **derivation** formed by adding the affix (here the suffix ·*ly*) to the base *quick*; finally, the lexical base *pencil* used as a verb (for instance in the phrase *I will pencil you in*) is a **conversion** of the lexical base *pencil* used as a noun.

3.4 Definitions

There are three ways to define concepts in grammar: 'notional', 'general', and 'language-particular' definitions. We can use any of these approaches to define concepts such as *noun*, *verb*, *preterite*, *imperative*, etc.

Language-particular definitions (those used primarily in CGEL) are those than focus on the grammatical properties (specific to the proposed grammar of the English language) one can use to distinguish one concept from another. For example, part of the definition of *noun* is that, 'nouns prototypically inflect for number (singular vs plural) and for case (plan vs genitive)'.

This is to be contrasted with traditional **notional definitions** that attempt to define grammatical concepts in terms of the meaning of the concept being defined. A typical notational definition of noun would be 'the name of a person, place, or thing'. This is an unsatisfactory definition as there are many abstract nouns, such as *rejection* in English that do not obviously deserve to belong in the category of 'things' any more than do related verbs such as *rejected*.

None the less, something like the notional definition is of value in a context in which we are interested in languages in general. Certainly the grammatical properties of nouns in French are not the same as those in English. It is therefore sensible to give **general definitions**, based in meaning, that apply in any language, as long as the category is grammatically distinct from others.

For example, a general definition of noun might be, 'a grammatically distinct category of lexemes of which the morphologically most elementary members characteristically denote types of physical object'.

The idea is that the general definition of a given concept should make sense for any given language as long as there is, for that language, a language-specific definition that is

based purely on grammatical properties (and hence the category is 'grammatical distinct' from other categories).

Thus there are many examples of grammatical case that have no meaning in English. For instance, the dative case is meant (by general definition) to mark the indirect object. But though we have indirect objects in English, they are not inflected. For instance, it would be meaningless to say that *the dog* is in the dative case in *I gave the bone to the dog* as *the dog* is not marked as being an indirect object.

3.5 Semantics and pragmatics

TODO: discuss:

1. The division between semantics and pragmatics.

2. Concepts of semantics: truth conditions, entailment, sentences, propositions, illocutionary meaning, conventional implicature.

3. Concepts of pragmatics: conversational implicature, presupposition.