Glossing of descriptive terms, and how to read a grammar

Jinyuan Wu

July 11, 2022

1 Theoretical orientation

The contemporary generative syntax is usually in a lexical-decomposition manner: that is, sub-word and even sub-morpheme features are the basic units of morphosyntactic operations. The features are fed into Merge to construct a binary tree with possible multi-domains (or a binary tree with chains recording copying or Internal Merge), and then the tree is linearized and post-syntactic operations are applied so features are assembled together to form words. This approach is not appropriate for language documentation, the latter requiring a more surface-oriented approach.

There have already been several relatively stable descriptive frameworks. Modern descriptive grammars of underdocumented languages are usually carried out within the framework of Basic Linguistic Theory (BLT), which is theorized in Dixon (2009, 2010, 2012), and, according to Dixon, deviates striking from the bond-to-fail generative approach. The BLT approach assumes no fine-grained constituency hierarchy, represents dependency relations not by constituency structure, and views constituent order as mostly driven by pragmatic reasons. In a word, BLT is the lay version of functional syntax à la Simon Dik.

On the other hand, already well-known languages are often described in the framework proposed in The Cambridge Grammar of the English Language (CGEL) (Huddleston and Pullum, 2002). This framework is also taken in Abeillé and Godard (2021), and its varieties are used in Demonte (1999); Muñoz et al. (2000); Bosque (1999). The CGEL approach is more informed by generative syntax: it maintains a largely binary-branching analysis, and dependency relations are still highly bounded to the surface-oriented constituency tree, though certain dependency relations that involve movements that are hard to find by looking at the surface form require special treatment: and thus the subject in a clause is recognized as a complement of the verb, but since it is outside the verb phrase (VP) and the A-movement to SpecTP is not recognizable in a surface-oriented, a new syntactic function label *external complement* is invented to cover the relation between the subject and the verb. In a word, the CGEL approach is the surface-oriented version of the mutual consensus of Minimalism (and GB) and HPSG.

Though the BLT approach, and the contemporary generative approach, or the Minimalist one, differ in several important aspects, the two frameworks roughly describe the same grammatical complexity class: I will show this in the rest of this note. For abstract discussion, see chap. 2 in my notes about Chinese syntax. The take-home message is CGEL may be viewed as the surface-oriented dual theory of the constituency structure of Minimalism, while BLT may be viewed as the surface-oriented dual theory of the dependency-based formalism of Minimalism. The relation between the three is visualized in Fig. 1.

Beside the discrepancy between the BLT and the CGEL approaches, there are still further divergences within the BLT approach. So here I list some dimensions of divergence:

- Is the theory purely lexicalist, or are there syntactic templates?[1] Though in a quick glance, it seems the lexicalist approach agrees with the Minimalist syntax while the templatic approach agrees with the constructionism, things are not that simple: remember, a Minimalist syntax runs on features which are not directly visible, and words and morphemes are just quirky

---

[1] Sometimes the term *lexicalist* means the syntax works on words and not sub-word units. This is not the meaning intended here. The meaning intended here by *lexicalist* is "all grammatical rules can be reduced to how to use certain lexicon entries (lexical or functional)", which may be words or morphemes or features. In other words, a lexicalist theory has no or few "global" phrase structure rules, as opposed to early generative grammars. This usage of the term *lexicalist* is attested in Matchin and Hickok (2020).
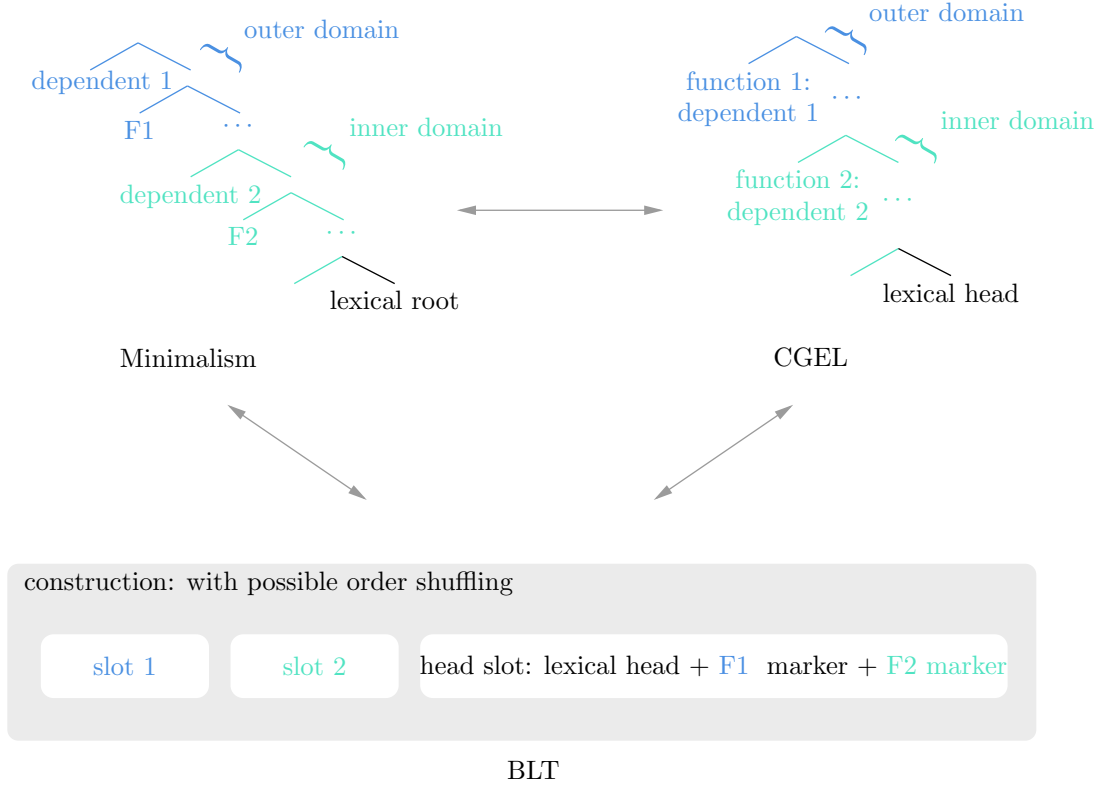
Figure 1: The generative formalism and two surface-oriented formalisms used in language description. The hierarchy information of the Minimalist formalism lost when deriving the surface-oriented formalisms is remended by grammatical functions like *indirect complement* and *function fusion* in the CGEL formalism, and by grammatical dependency relations like "slot 1 is the syntactic topic" and "the dependency relation between the lexical head and the F1 slot is coded by the F1 marker" in the BLT formalism.

reflections of them. The corresponding surface-oriented version of a Minimalist syntax with lots of features that are used to guide the syntactic derivation (e.g. the EPP feature), then, inevitably contains syntactic templates that are hard to place under any lexicon entry. The Cinque hierarchy of clause structure, for example, contains tons of invisible functional heads, and once we "integrate out" these functional heads, the resulting grammar has a clause template. The linguist has to consider whether to introduce a chapter named "the structure of noun phrases" or a chapter named "the clausal structure".

- How is morphology dealt? This parameter has strong association with the previous parameter, since there is no clear distinction between a morpheme and a word. In morphology the lexicalist extreme is the Item-and-Arrangement approach, while the templatic extreme is the Word-and-Paradigm approach. The Item-and-Process approach is somehow in the middle, maybe in a position closer to the former and further from the latter. What brings in more complexity in morphology is there are post-syntactic operations: even when the features do spellout into morphemes, the Distributed Morphology-style post-syntactic operations blur the correspondence between features and morphemes, and hence the idea that words are built up by morphemes does not lead to any constraints on the form of the word, raising doubts about whether in a surface-oriented analysis, morphemes are of any theoretical significance at all (Anderson, 2017). The linguist needs to pick up a specific way to show how words are built up.

- How are grammatical relations (in other words, dependency relations) introduced? Together with morphemes that bear them, or words, or constituents, or with separate chapters and sections? This parameter has certain correlation with the top-down/bottom-up parameter, because in a top-down analysis, the grammatical functions of constituents in a larger construction are obviously introduced before what fill the constituent slots are discussed. On the other hand, a bottom-up grammar tends to introduce grammatical relations when discussing the smallest unit that bear them, for example talking about the case marking of various complements in the noun morphology chapter.

- What is the relation between a phrase and words contained in it? What is the head? What are the complements? What are the modifiers? In Minimalist syntax, all functional categories serve as heads, but lexical categories are never heads. This may appear strange but has underlying consistency (see § 2.1.2.4 in my notes about Chinese syntax). This approach, however, is not acceptable for a surface-oriented grammar, and here another concept – what determines the "overall" property of a constituent – is accepted as the standard to decide what is the head. Thus a nP and a DP are all headed by the central noun in the surface-oriented analysis, because both of them are built surrounding the core noun stem, and since the core noun stem is phonetically realized as the central noun – a lexical word – the latter is recognized as the head. Disagreements then arise when whether a word is functional or lexical is not that certain. Should the preposition be considered as a head? The preposition in a peripheral argument may be seen as the marker of a syntactic case system (so in the generative analysis, we have PP and CaseP), and under this analysis, the preposition is not a head. But in many languages like English, the preposition category has certain predicative properties, making it appear like a lexical category, and then it seems a noun phrase (NP) with a preposition is no longer a NP – it is a preposition phrase (PP) headed by the preposition.

- Are there fine-grained constituency structures, or are there just noun phrases and clauses? Some grammars, like the CGEL, posits an anatomy of NPs with the following functional domains: head noun – nominal – minimal NP with a determiner – external modifiers. Others just list possible NP dependents or clausal dependents, without discussing which is closer to the head. If the latter approach is taken, the linguist has to introduce effects due to the relative position of constituents in another way, like "the O argument in ergative languages is more topic-like". The main reason to take the latter approach – which is the approach advocated in BLT – is only NPs and clauses have complete semantic significance. See BLT § 1.11, (33) and (34): Dixon does not like the binary-branching (Minimalist) approach (33), because it

does not illustrate the fact that the functional words are different from lexical ones. But this is more a problem of terminology: the term *phrase* in BLT corresponds to a maximal domain like DP or CP in generative syntax, while a generative *phrase* – like vP or AdvMannerP – corresponds to a grammatical construction in BLT.

- How is constituent order (often called *word order*) introduced? Is there a separate chapter devoted to constituent order? Constituent order can be understood as a manifestation of constituent hierarchy, while in more functionalist approaches, it is understood as a method parallel to morphological marking that marks the constituent positions in a larger construction. Note that the second claim does not go against the first one: certain features are indeed reflected by the surface constituent order in generative syntax, Certain ideas in the first account that do not involve features (e.g. Antisymmetry) cannot be translated transparently back into the second approach, though, but they can be framed in the second approach as "the human language faculty just rejects certain constituent orders anyway".[2]

- Top-down (i.e. structuralist partition-based), or bottom-up (i.e. based on the usage of smaller units)? In PSGs there is a clear correspondence between the two, but for actual language documentation things are often complicated: a top-down grammar is awkward to write because the author has to enumerate all possible configurations in a construction to fully characterize it ("a clause is either coordination of clauses or a subject-predicate construction" – oh no, supplementation and pre-nucleus constructions are forgotten), while a bottom-up grammar is awkward to read because the reader has to infer all possible configurations in a construction ("the verb is the prototypical content of the predicate slot" – any other possibilities? Nobody knows). This parameter is in principle orthogonal to the parameter about how constituent order is introduced, but a bottom-up grammar without a chapter (or several chapters) devoted to constituent order will be extremely hard to read: the reader may find a sentence like "the object follows the verb" in the chapter about verbs. Alright, can an adverb intervenes between the verb and the object? No answer.

- Whether a set of canonical constructions is established. Viewing non-canonical constructions as transformed from canonical ones (or by adjunction, etc.) is a powerful descriptive tool, but it is often the case that certain constructions that are uncontroversially deemed non-canonical do not have a canonical counterpart. This is one of the reason transformational rules are finally abandoned in generative syntax. Transformational rules (and adjunction, etc.) are still handy when doing description, though: no one wants to read a grammar that treats positive clauses and negative clause in the same way.

- How corpus examples are given – fully bracketed and labeled, or represented as is? In the first case, an example is a demonstration of the grammatical *rules* generalized in the grammar, while in the second case, an example is just there as a piece of observation. This parameter has relation with the top-down or bottom-up parameter, but the relation is not absolute: in general, the more bottom-up a grammar is, the more observation-based it is, i.e. the examples are provided as is without much tree diagrams or bracketing expressions. But it is of course possible to have a top-down grammar which does not make much generalization about grammatical rules that generate grammatical utterances and exclude ungrammatical ones. Indeed, this is exact the case for the structuralist Immediate Constituent Analysis: binary partition of corpus data is made, but no glossing about what the tree diagrams obtained is provided, and there is no generalization about what is permitted and what is not. An example of grammars of this kind is Chao (1965). The correlation between the bottom-up narrative order and the observation-based approach is likely due to it is the easiest way to document something about a language without mistakes.

---

[2]One controversy here is the generative feature-driven constituent order often involves movement, while functionalists accept constituent order variations "as they are". This controversy is false, because for many generative linguists, movements can be unmarked, and what movement means is simply dual syntactic function of a constituent or the imperfect relation between constituent order and dependency relations (e.g. cross-serial dependencies).

Both BLT and CGEL have strong tendencies in the values of some (though not all) parameters listed above. The parameters that are largely fixed by the overall descriptive framework include whether fine-grained constituency hierarchies appear, how is constituent order introduced, and the definition of *head*. It is of course possible to mix the BLT and the CGEL approaches: Friesen (2017), for example, despite being a largely BLT-like grammar, uses the term *verb phrase* in the way of CGEL, while the parts about NPs are typically BLT-like. On the other hand, Muñoz et al. (2000), despite it employment of CGEL-like terms, does not emphasizes on tree diagrams and constituency structure, nor its reflection on the constituent order, and is much more BLT-like than the CGEL itself or the great French grammar (Abeillé and Godard, 2021). Another example of mixing of the CGEL approach and the BLT approach can be found in CGEL itself: ditransitive clauses can be analyzed in Minimalism, but the analysis inevitably involves lots of movements to make the dependency relations correct, so CGEL just gives up binary branching in exchange of a much clearer surface-oriented analysis, where the syntactic properties of the direct and indirect objects (two grammatical functions not found in monotransitive clauses) are manually stipulated. This is just how BLT works when the constituency relations are complicated.

Choosing between the approaches in CGEL and BLT and the rest of the parameters are to be decided by the grammarian, and some best practices are introduced in § 2. In principle, the above parameters are free to choose. In practice, they have to be fine-tuned or otherwise the grammar will be hard to read.

2   (Not so) best practices of grammar writing

2.1   What influences the organization of the grammar

Parameters in § 1 have to be fine-tuned according to the following factors:

- The properties of the language. If a linguist unfortunately decides to write a Chinese grammar in a bottom-up manner in which a grammatical relation is introduced in the chapter about the lexical category about its head, a reader will soon be stuck in questions like what are the possible linear order between object(s), directional complement, and aspectual markers. On the other hand, it is okay – and even desirable to write a Latin grammar in this way, because Latin is much more free-order than Chinese and grammatical relations are mostly marked by morphology.

- Whether the language is already well-known. In § 1 we have already seen that comprehensive grammars of well-known European languages are usually CGEL-like, while newly documented languages are usually captured by BLT-like grammars. This is again expected, because in-depth partition of constituents is never the priority when describing a newly encountered language.

2.2   Organization of chapters

2.3   Terminology

2.3.1   Form-based grammatical relation terms

Some people tend to use the lexical or phrasal category that prototypically fills a grammatical function slot as the name of that grammatical function. Here is a list of relevant terms:

- *Adverbial*, which is actually peripheral argument (Box 3.1).

- *Serial verb construction*, which is actually serial predicator construction.

- *Verb complex*, which is actually predicate (in the BLT sense) or the predicate minus complements (in the CGEL sense).

3 Top-down partition of the clause structure

3.1 A sketch of the constituency tree

This section gives a purely form-based analysis of clause structure. Just like when discussing morphology, we often first show possible morphological devices and then discuss what grammatical categories are marked by these devices, in this section, I first discuss how to give a constituent analysis of a clause, and then in other sections about how to interpret the constituency tree obtained.

3.1.1 A clause is built up by one or more nuclei with certain syntactic processes

The top-level partition of a clause is given as the follows:

(1) A **clause** is

- the coordination of two clauses (§ 7.1), which may involve ellipsis in and/or movement out of the conjuncts, or
- a clause with supplementation (§ 8), or
- a clause without the two.

(2) A clause without coordination or supplementation is

- a clause with pre- or post-nucleus constructions (the residue of the nucleus clause undergoing relevant syntactic processes is named the **nucleus**), like the English subject-auxiliary verb inversion or *wh*-movement, or
- a nucleus clause (see (3)).

Note that the distinction between coordination and adjunct clause construction (a type of subordination) may be not so clear for some languages, for example Latin (see my notes about Latin, § 2.7). Also, there is no strict application order between coordination, supplementation, and pre- and post-nucleus constructions: in the English question *on that particular day – I mean the day when the unfortunate incident happened – did you pass that site or hear anything usual in that direction*, first a coordination construction is used, followed by a subject-auxiliary verb inversion (a pre-nucleus construction) and then supplementation and finally another pre-nucleus construction (topicalization of the time adjunct). Another remark here is the syntactic processes from nucleus clauses to more complicated ones may only work for certain inputs: in English, for example, the supplementation *not even …* is only possible for a clause in negative voice. A final remark is that finding the boundary of the nucleus requires testing the transformational properties of each clausal dependents. The constituency tree obtained by immediate constituent analysis is without labels, and it is impossible to decide, say, whether a constituent is a topic (a prenucleus construction) or a subject (a dependent within the nucleus). This fact leads some Chinese grammarians to abandon the distinction between topic and subject. Here is not the best space to have in-depth discussion for the definition of the nucleus. A working definition is given in § 3.1.2, but it involves terms like *complement* and *adjunct*, which cannot be defined by the label-free constituency tree obtained by clause partition.

3.1.2 Clausal dependents in the nucleus

Now it is time to define the nucleus clause:

(3) A **nucleus clause** is

- a minimal nucleus clause, or
- a nucleus clause with adjunction.

(4) A **minimal nucleus clause** is a complex of

- the **predicator**, prototypically a verb but with possible alternatives, possibly marked for grammatical categories involved in the clause structure, and

6

- one or more visible or invisible **complements**, and
- possible functional words marking clausal grammatical categories

or it is a serial verb construction (§ 3.3).

Here **adjunction** means adding **adjuncts** into the tree structure, in the manner in Tree-adjoining grammar (TAG). This is the surface-oriented counterpart of optional projections in Cinque hierarchies. Adjuncts are contrasted with complements, the latter being somehow closer to the predicator, but not necessarily obligatory. There are several tests to find whether something is a complement or an adjunct (see CGEL § 4.1.2, for example), but the distinction is usually quite blurred and language-specific (§ 2.1.3.4 in my notes about Chinese syntax). The complement-adjunct distinction is usually hard to test simply by pure constituent analysis, because an adjunct is not necessarily higher than all complements. The distinction is not what can be studied here – it has to be delayed to § 3.1.4.

The term *adjunct* used in this note means clausal modifier. *Adjunct*, in generative syntax, means optional non-head components of any projection, though nowadays, especially in the Syntactic Cartography program, it is often assume that there is no adjoin operation beside the usual Merge, and so-called adjuncts are specifiers of certain optional functional heads, and hence the term *adjunct* loses its structural significance. Many descriptive grammars, like (Quirk, 2010), use the term *adverbial* for the term *adjunct*. A third name used for adjuncts are *peripheral argument* in BLT.

The term *complement* may sometimes be used to denote specifically *complement clauses*. In BLT, the term *complement* is usually replaced by *core arguments*.

CGEL insists on a strict form-function distinction and hence the term *argument* is reserved for semantics. BLT, on the other hand, emphasizes on the semantic basis of syntax, and so the term *argument* is used. But here comes a subtle difference between BLT's standard of clausal dependents and CGEL's: certain constituents, like the direction complement in Mandarin Chinese (see § 10.2 in my notes about Chinese syntax), are definitely complements under the standard of CGEL, but are definitely not arguments, and hence they are not recognized as clausal complements in BLT – they are thus recognized as a part of the BLT predicate (see Box 3.2).

3.1.3 Pre- and post-nucleus constructions not well defined for free-order languages

It should be noted that for languages with a relatively free constituent order, it is almost impossible to find a neutral order, and hence pre-nucleus and post-nucleus constructions cannot be well-defined, let along the fact that some linguists posit so-called in-VP scrambling and the pronominal argument construction for radical non-configurational languages where argument NPs are actually adjuncts, which are by no means pre- or post-nucleus constructions but nonetheless induces changes in the constituent order. In this case, (2) and (3) should be merged together, and notions like pre- and post-nucleus constructions are to be replaced by discussions on the relation between constituent order and semantic and pragmatic information.

3.1.4 Classification of clausal dependents

In the above discussion I intentionally avoid mentioning subject and object or more generally, any complement and adjunct types, because as is discussed in the end of § 3.1.1 and in § 3.1.2, by staring at the surface form, it is impossible to define these terms. The constituent analysis of the surface syntax gives us a constituency tree without labels. and in this section, I discuss how to add labels to the nodes of the tree.

Trivially, the label of a node can be assigned as the category of the node. Thus we have $[_S [I]_{Pron} [_{VP} [like]_V [fruits]_{NP}]]$. But of course *I* and *fruits* occupy different syntactic positions: from folk grammar we know the first is the subject while the second is the object. Can this distinction be seen from their structural positions? It is true that in generative syntax, complement positions can be distinguished purely in structural terms (SpecTP, Spec$v$P, etc.), but in the surface-oriented analysis the relevant functional heads are all invisible, and inevitably there are occasions when structural terms are occasionally insufficient to decide the role of a clausal dependent. A grammarian may want

to define the *subject* as "the most external clausal dependent in immediate constituent analysis", and this leads to the ridiculous conclusion that in *quietly, he entered the room*, the adverb *quietly* is the subject.

This example is a vivid example of the importance of *functional* labels in the constituency tree.[3] Another illustration can be found in Fig. 1 in my notes about CGEL: the information contained in the functional projections in the left tree is displayed in the functional labels like *predicator*, *agent*, *patient*, etc. in the right one. In the above example of the definition of the subject, "the most external clausal dependent in the nucleus" seems to work well. But *what* is the nucleus? This question is raised at the end of § 3.1 but has never been answered definitely. One may want to define it in semantic terms: "a nucleus clause contains all constituents that are necessary to complete the meaning of the verb". But isn't *quietly he entered the room* such a clause?

The conclusion is that in order to find complement positions, adjunct positions, the definition of *nucleus* – in a word, functional labels – in a language, syntactic tests based on transformational behaviors (or to be precise, baed on comparison between regularly related constructions) is needed. The grammarian should expect lots of cross-linguistic variation of available functional labels, because they are reflections of the underlying feature structure. In a purported language without the EPP feature, SpecTP may not be not of much significance, and hence the label *subject* is of little use. An example of how to find complement types is Fig. 11 in my notes about CGEL. By six syntactic standards, five object-like complement types are identified, and four of them are recognized as objects with one being kicked out of the object family.

3.1.5 Alignment

Recognizing complement types is mostly about syntax. Semantics is also involved, of course, because otherwise it is impossible to decide whether one clause is transformed from another. But this is not what people already with knowledge on traditional grammar expect. What do semantic roles of complements and adjuncts do in complement type recognizing? They are important factors in determining complement and adjunct types, but they are by no means the decisive ones. Passivization is a well-known counterexample of the uniform matching between semantic roles and complement types. A locative role may be realized as a complement or an adjunct. The defective – yet still largely regular – mapping between semantic roles and complement or adjunct types is therefore worth separate treatment. This is what known as **alignment**.

There are usually much fewer complement types than semantic roles. The detailed semantic classification of verbs – and their complements – is quite messy. Table 3.1 in BLT lists the semantic classification of the most frequent verbs in English. There are 16 semantic roles mentioned, and if the linguist just lists his/her testing results of the syntactic behavior of each of them, the reader will be driven mad! A well-organized description of the alignment of a language therefore is based on the mapping relation between complement and adjunct roles and coarse-grained semantic roles. A priori no one knows how to coarse-grain semantic roles so that the resulting "macro-roles" make sense both syntactically and semantically. After periods of investigation, several macro-roles have already been identified. The most important three are S, A, and O[4]. S is the only argument in the intransitive clause. A and O are the more agentive argument and the more patientive argument in the transitive clause, respectively. Other argument labels include E (whatever that occurs together in addition to a SV or AVO construction – see § 4.5.2 in my notes about CGEL), G (the goal-like argument in a ditransitive construction with a meaning of "transferring" – see § 4.5.3 in my notes about CGEL), and T (the theme-like argument).

If somehow S and A are realized by the same complement type – which is commonly named the *subject* – then the language is **nominative-accusative** or simply **accusative**. If, however, S and O are realized by the same complement type, then the language is **ergative-absolutive** or simply **ergative**.

The requirement for them to have largely uniform syntactic behavior

---

[3] Here *functional* means syntactic function and not pragmatic function.

[4] Many denote it as P, to be consistent with the A argument (agentive/patientive). In this note I use O to be consistent with Dixon.

> **Box 3.1: Terminology: adjunct, adverbial, complement, arguments**
>
> This note keeps all the four terms: *complement, adjunct, core argument, peripheral argument.*

3.2 The inner structure of the verb-complement construction: syntactic alignment

3.2.1 The subject-predicate analysis of the verb-complement construction for syntactically accusative languages

(4) is a flat-tree analysis, but there are several evidences suggesting a fine-grained hierarchy is useful even for surface-oriented analysis. For accusative languages, the S and A arguments are and hence are identified as the *subject*, and we have the following facts:

- The subject is much easier to be extracted out of the nucleus, which can be explained by the theory that it is somehow higher and movement operations are localized.

- The quantifiers of the subject and internal complements, explicit or implicit, demonstrate a stable scope hierarchy: the scope of the subject quantifier is always larger. When talking about a charity organization, one may say *every woman helps three boys.* Here, the subject is bounded by $\forall$ and the object is bounded by 'there exists three ...', and $\forall > three$ and *$three > \forall$: the meaning of the sentence aforementioned is 'for each woman, there are three boys that she helps, but I do not know who they are, and possibly the boys Sarah helps are not the boys Lily helps'. After a seemingly trivial passivization, we get *three boys are helped by every woman*, which means 'there are three boys – I don't know who, but anyway there are three – who are helped by every woman in our organization', and we have *three* $> \forall$ and *$\forall > three$. If we assume the semantics is related to the syntactic structure at least partially, then this is a piece of evidence that the subject is higher in the syntactic tree, no matter what its semantic role is.

- If the subject is indefinite, then it is by default bounded by $\forall$, TODO: really???

  Some notes about BLT chap. 13, Appendix 1: TODO: S argument and A argument are by default bounded by $\forall$, while O is bounded by $\exists$ – is this cross-linguistically correct? This also explains why verb-object incorporation is frequent: *a cat kills some animals = a cat kills.* It seems the only argument – be it peripheral or core – that is by default bounded by $\forall$ is S in intransitive clauses and A in transitive clauses (which may be seen as a double check). What's the counterpart in syntactic ergative languages?

- Verb-argument incorporation, nominalization, etc. (for example compare *solve problem* and *problem solving*) usually happens between the verb and the internal complement(s), not between the verb and the subject.[5]

- If there is something looking like reflexive pronouns, then it usually follows the Government and Binding scheme, and using this as a test, the subject is always predicted to occupy a higher position.

The list can go on and on, and hence it is useful to divide the nucleus clause into the subject and the predicate:

(5) A nucleus clause is made up by an **external complement**, often named the **subject**, and a predicate.

(6) A **predicate** is either a predicate without adjunction, which may be

---

[5]Grammaticalization of a span is also in principle possible, so incorporation between the verb and the subject may still rarely occur. But note that operation on span is usually seen for functional hierarchies, in which what are spellout as a single word are highly lightweight functional heads, not more substantial lexical categories.

- a predicator-complement construction, or

- a serial verb construction,

  or a predicate after adjunction, or a predicate after syntactic processes marking clausal grammatical categories like negation, modality, etc.

(7) A predicator-complement construction consists of a predicator (which is the head of the predicate and the clause) and its **internal complements**.

The above rules replace (3) and (4). These are the rules used in CGEL, and actually also the Chinese school grammar (see § 6.1 in my notes about Chinese syntax). CGEL does not acknowledge the role of verb complex, which is in principle correct, because English is already highly analytic and rigid-order, and most information about dependency relations can be reconstructed from the surface-oriented constituency analysis. Therefore, the term *predicate* is used as in traditional grammar: it means the nucleus minus the subject, and its status as a constituent summarizes the above listed facts.

3.2.2 The flat-tree analysis of the verb-complement construction

It should be noted the above concept of *predicate* does not always correspond to an uncontroversially constituent in the surface structure: in a VSO language, for example, the predicate is discontinuous. This urges some to accept a flat-tree approach to describe the nucleus predicate.

To see another motivation of the flat-tree approach in surface-oriented description, consider the following facts. There are some controversies arising from the "what is the head" parameter in § 1. Some verbs are auxiliary verbs. In the clause *I should do this*, what is the predicator? Here we are in the same delimma as the one concerning "preposition phrase". An analysis in which the predicator is *should* will face the criticism that functional words are never heads in a surface-oriented analysis, or otherwise, in order to be self-consistent, its bound morpheme counterparts should also be regarded as heads, which falls back to the generative functional head analysis.

The above two motivations urge us to take the analysis in which the main verb *do* is the predicator. This approach usually occurs with the flat-tree approach, or otherwise *should* is analyzed as a clausal dependent similar to the determiner in a NP, which is acceptable in the structuralist analysis of Chinese non-argument complements (see chap. 10 in my notes about Chinese syntax) but is not prevalent outside the Chinese grammar community. The nucleus minus arguments (core and peripheral) is named *verb phrase* in BLT, while in CGEL the term *verb phrase* means the verb plus its internal complements, which is the form of the predicate. The BLT *verb phrase* is the hierarchy *span* (as in *span spellout*, or the head slot in the BLT part in Fig. 1) of the CGEL VP shell. To avoid conflict, the term *verb complex* may be used to denote the verb phrase in the BLT sense (Hockett, 1948; Friesen, 2017; Wilbur, 2014). Now since there is no need to do fine grained partition of the nucleus, the term *predicate* can be assigned to something else and the commonly accepted practice is to use it to denote the verb complex. In this way, (5), (6), and (7) are replaced by the following:

(8) A nucleus clause is made by a predicate (with a different meaning with the *predicate* in (6)) and core and peripheral arguments. Rearrangement of constituent order may be necessary.

(9) A **predicate** may be a simple one or a serial verb construction (§ 3.3).

(10) A simple predicate consists of a head (prototypically a lexical verb) and possible grammatical category markers (including auxiliary verbs).

This approach is not without doubt: the analysis that the main verb *do* is the main verb also faces a problem of non-consistency: the boundary between auxiliary verbs and lexical catenative verbs is somehow vague in some languages, and in this case, in somewhere in the grammaticalization process the head status suddenly flip from verb to another. Also, the problem of discontinuous constituent (the verb complex) still exists, which can also be seen from Fig. 1. But this is the practice accepted in BLT and most of grammars sticks to this paradigm, so I introduce it here.
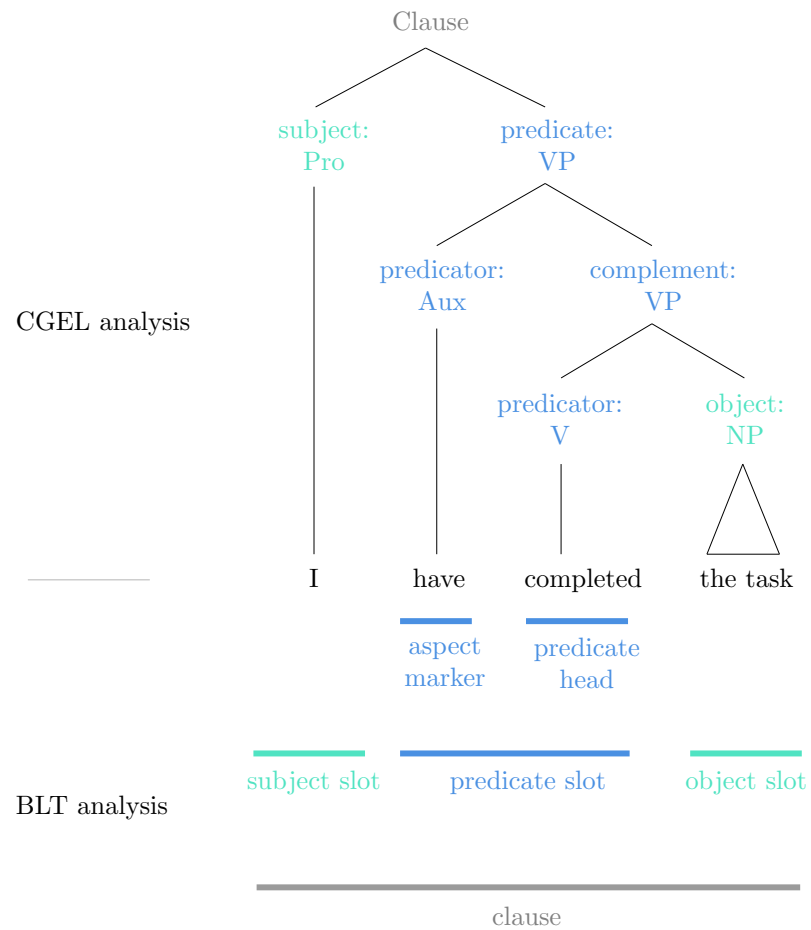
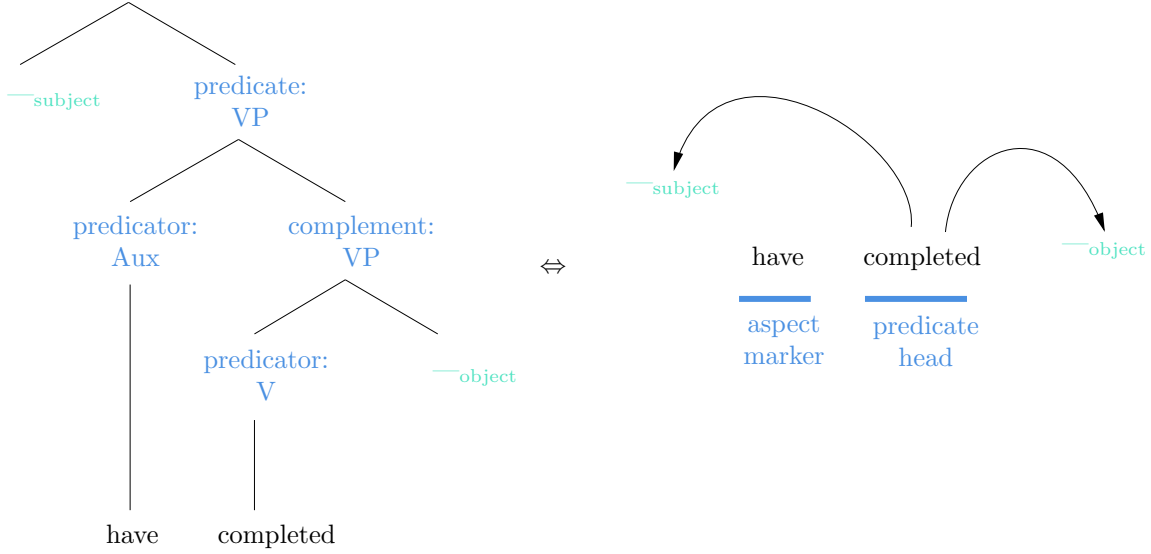Figure 2: Comparison between a BLT analysis and a CGEL analysis

Figure 3: The structure of a simplest BLT predicate

3.2.3 Switching between the two

Finally I discuss how to translate between the two approaches. There are only two key points: in BLT, most heavy lifting jobs are done by dependency relations within a large, non-branchable constituent, not constituency relations, and what is a constituent depends heavily on semantics. Consider Fig. 2, which illustrates the divergences between CGEL and BLT over a simple predicator-complement construction. Here is a comparison between the two analyses:

- The argument positions in both approaches are labeled by explicit functional labels. The necessity has been argued in § 3.1.4: in Minimalism, the argument positions can be decided purely in terms of the structure: the subject is SpecTP, and object is SpecTransP, etc. But in the surface-oriented formalisms information contained in the constituent structure is not enough to tell us the role of arguments, and hence arguments are explicitly labeled via notions like "subject:", "object:", "subject slot" and "object slot".

- The binding relation between the subject and the object is realized in BLT as "the subject tends to somehow control the object cross-linguistically". This reflects the dependency relation-based nature of BLT: the fact that the subject is somehow "higher" than the object is not reflected in the constituent structure, but rather by a dependency relation placed in the clause template.

- The fact that *have completed* is a span in the TP projection is reflected in the CGEL analysis also by constituency structure, but *have completed* is recognized as a constituent in the BLT approach. This reflects the semantic-informed constituent standard in BLT: as can be found in the CGEL part of Fig. 2, in the clause structure we have a VP *have completed the task*, but it is neither a clause nor a NP, so it is not a constituent in BLT, but once the NP *the task* is stripped away, the purely verbal skeleton *have completed* – which is a span and colored as blue – is of semantic significance: it describes the event going on, and the internal gap for the object as well as the external complement (i.e. subject) gap are interpreted as valency (here the dependency-based nature of BLT is also reflected). So *have completed*, a span in CGEL, gets recognized a constituent in BLT and it fills the predicate slot. This is illustrated in Fig. 2.

Similarly, the generative constituency relation-based illustration of serial verb construction in e.g. Chen (2016) is replaced the dependency relation-based syntactic process to insert several verbs into a single predicate and alternate the dependency relation between the verbs and the arguments in

12

BLT. Fig. 4 is an illustration of the correspondence between the analysis in Chen (2016) and a BLT analysis. The phenomenon discussed in the diagram is the direction complement construction or direction compounding construction (see § 10.2 in my notes about Chinese syntax), in which a word indicating spacial movement (picked from a closed category) – which is the direction complement or direction compound – is merged together with a TransP, and the object – in this case *tāng* 'soup' – in the TransP also becomes a complement of the direction complement, and by head movement, the direction complement is attached to the main verb – *sòng* 'send' here – and hence a compounding verb *sòng laí* is formed, which precedes the object. Now we try to translate the above derivational process to a BLT account:

- The fact that *tāng* is both the complement of the main verb *sòng* and the direction complement *laí* (by appearing in two specifier positions) is reflected by the two dependency relations on the right side.

- The fact that *sòng laí* is the span spellout of the verbal projections minus the object is reflected by putting the two words into the predicate slot. The constituent order information is reflected in a flat-tree way.

Here I intentionally skip the CGEL approach because the complexity of serial verb construction is already comparable with the functional projection that gives rise to the *have been being done* hierarchy (Ramchand and Svenonius, 2014), and the surface-oriented CGEL analysis will not be substantially different from the BLT analysis shown in Fig. 4.

There are adjustable parameters in both the CGEL approach and the BLT approach. One

> **Box 3.2: Terminology: predicate, predicator, and verb complex**
>
> The predicator is the head of the predicate, which is always filled by a verb in English.
> In flat-tree approach grammars, i.e. BLT, the

> **Box 3.3: Summary of the constitueny tree of clause structure**
>

3.3   Serial verb construction

3.4   Grammatical systems and categories in the clause

4   Top-down partition of the noun phrase

5   Lexical categories

5.1   The noun category

5.2   The verb category

5.2.1   What is recorded in the dictionary entry of a verb

Here is a list of what needs to be described in the dictionary entry of a verb if the dictionary is expected to provide full information instructing how to build a sentence from words:

- Grammatical categories marked on the verb and

Figure 4: An example of serial verb construction: *send come soup* is the literal translation of the Chinese *sòng laí tāng*, which means 'send soup here (= send soup and make it come here)'. Based on (74) in Chen (2016), which adjustment to make it compatible with Distributed Morphology.

References

Anne Abeillé and Danièle Godard. *La grande grammaire du français.* Éditions Actes Sud, 2021.

Stephen R Anderson. Words and paradigms: Peter h. matthews and the development of morphological theory. *Transactions of the Philological Society*, 115(1):1–13, 2017.

Ignacio Bosque. *Gramática descriptiva de la lengua española: Entre la oración y el discurso. Morfología*, volume 3. Espasa, 1999.

Yuen Ren Chao. *A grammar of spoken Chinese.* ERIC, 1965.

Zhishuang Chen. *Mandarin directional serial verb constructions: A constructionist approach.* PhD thesis, University of York, 2016.

Violeta Demonte. *Gramática descriptiva de la lengua española: Sintaxis básica de las clases de palabras*, volume 1. Espasa Calpe Mexicana, SA, 1999.

RM Dixon. Basic linguistic theory. volume 3: Further grammatical topics, 2012.

Robert MW Dixon. *Basic linguistic theory volume 1: Methodology*, volume 1. OUP Oxford, 2009.

Robert MW Dixon. *Basic linguistic theory volume 2: Grammatical topics*, volume 2. Oxford University Press on Demand, 2010.

Dianne Friesen. *A grammar of Moloko.* Number 3 in African Language Grammars and Dictionaries. Language Science Press, Berlin, 2017. doi: 10.5281/zenodo.824016.

Charles F Hockett. Potawatomi iii: The verb complex. *International journal of American linguistics*, 14(3):139–149, 1948.

Rodney Huddleston and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language.* Cambridge University Press, 2002. doi: 10.1017/9781316423530.

William Matchin and Gregory Hickok. The cortical organization of syntax. *Cerebral Cortex*, 30(3): 1481–1498, 2020.

Ignacio Bosque Muñoz, Violeta Demonte Barreto, Fernando Lázaro Carreter, and María Victoria Pavón Lucero. *Gramática descriptiva de la lengua española: Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales.* Espasa-Calpe, 2000.

Randolph Quirk. *A comprehensive grammar of the English language.* Pearson Education India, 2010.

Gillian Ramchand and Peter Svenonius. Deriving the functional hierarchy. *Language sciences*, 46: 152–174, 2014.

Joshua Wilbur. *A grammar of Pite Saami.* Number 5 in Studies in Diversity Linguistics. Language Science Press, Berlin, 2014. doi: 10.17169/langsci.b17.34.