Glossing of descriptive terms, and how to read a grammar

Jinyuan Wu

July 10, 2022

1 Theoretical orientation

Modern descriptive grammars are usually carried out within the framework of Basic Linguistic Theory (BLT) (Dixon, 2009, 2010, 2012), which, according to Dixon, deviates striking from the bond-to-fail generative approach, though the two frameworks roughly describe the same grammatical complexity class: I will show this in the below discussion. Indeed, a generative theory which relies on a fixed functional hierarchy may even be branded by the author as *constructionist* (Chen, 2016). The contemporary generative approach, or the Minimalist one, is based on features that are Merged and undergo certain morphophonological processes, and BLT may be viewed as the surface-oriented dual theory of it. (see chap. 2 in my notes about Chinese syntax)

There are still further divergences within the BLT approach. Here are 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 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

[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).

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 *n*P 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 The Cambridge Grammar of the English Language (CGEL) (Huddleston and Pullum, 2002), 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 *v*P 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

[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).

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.

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. 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.

There have already been several

2 Best practices of grammar writing

2.1 Organization of chapters

2.2 Terminology

2.2.1 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 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

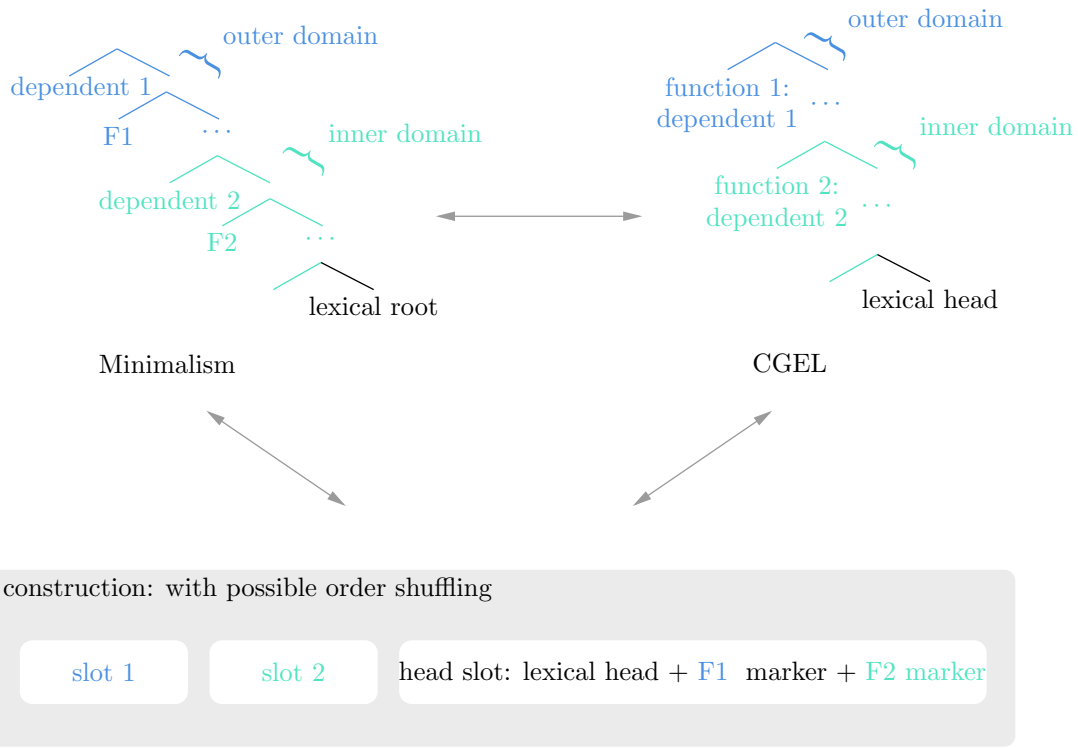

Figure 1: The generative formalism and two surface-oriented formalisms used in language description. The hierarchy information of the Minimalist formalism lost when deriving the surface-oriented formalisms is remended by grammatical functions like *indirect complement* and *function fusion* in the CGEL 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.

- 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.

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
- 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 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).

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 TODO: ref

Box 3.1: Terminology: adjunct, adverbial, complement, arguments

This note keeps all the four terms: *complement*, *adjunct*, *core argument*, *peripheral argument*.

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 The inner structure of the nuclues clause 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.[3]
- 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

- 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.

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:

[3] 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.

(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.

Finally I discuss how to translate between the two approaches. In BLT, most heavy lifting jobs are done by dependency relations within a large, non-branchable constituent, not constituency relations. The constituency relation-based illustration of serial verb construction in e.g. Chen (2016) becomes 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 BLT.

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

Parameter: the flat-tree analysis in BLT means the term *predicate*

TODO: what is included in the verb complex? Auxiliary verb, serial verb construction, of course, but what about lexical catenative verbs?

Box 3.3: Summary of the constitueny tree of clause structure

3.2 The simplest verb-complement construction and alignment of core arguments

Alignment means

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

5.2.2 Semantic classification

5.3 Distinguishing nouns and verbs

5.4 The adjective category (or categories)

6 Morphology

7 Coordination

7.1 Clausal coordination

7.2 Coordination in NPs

8 Supplementation

References

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.

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

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.

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

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.