X-Bar Theory: A General Scheme for Phrase Structure Rules

Renjini Narendranath

Dept. of Computational Linguistics, Universität des Saarlandes
renjinin@coli.uni-sb.de

October 1, 2003

Abstract

This article provides an overview of the X-bar theory which is one of the fundamental components of the Government and Binding theory as proposed by Chomsky. Before explaining the theory formally, we introduce some of the basic terminologies and motivation behind the X-bar theory by considering the common patterns in the phrase structures and sentences. We also discuss the extension of X-bar theory to different languages.

1 Introduction

X-bar theory was developed in the 1970s by Chomsky and it is one of the fundamental components of Government and Binding theory (GB)[2]. GB is in a sense the immediate descendant of Transformational Grammar (TG). The idea behind TG is to transform related structures into each other. Transformations can for instance express the relationship between a declarative sentence and a question. The role of X-bar theory in GB is to characterize the commonalities in the phrase structures. What X-bar theory postulates is that, we can find similar patterns recurring in the structures of phrases. Hence by abstracting away the language specific aspects like word orders, (Eg: in English, subject comes first, then verb, and then objects) we can talk about language independent template of X-bar theory for sentences and phrase structures. There are many literatures that discuss the X-Bar theory extensively[2, 3, 4, 5]. X-bar theory also plays an important role in the formation of the structure of phrase rules in the Generalized Phrase Structure Grammar (GPSG) and Lexical Functional Grammar (LFG) [5].

In section 2, we discuss some of the basic terminologies that are needed for understanding the X-bar theory. In section 3, we try to abstract the common patterns that occur in phrase structures as well as in sentences. Using these patters, we first consider a special case of X-bar schema for verb phrases in English language. In section 4, we formalize the above findings to a general X-bar schema for all phrase structures in English. We also investigate a possible X-bar schema for sentences. We further extend the schema to other languages in section 5. In section 6, we look at the global picture of GB and place X-bar theory appropriately in it. In section 7, we briefly discuss some parser implementations based on X-bar theory.

2 Some Basic Terminologies

In this section, we discuss some of the basic terminologies related to phrases and its constituents.

- *Lexical Categories*: The basic (least decomposible) constituents of a phrase or a sentence. They constitute what is called the deep Structure of Phrase Structure Grammar. Examples: Nouns (abbreviation- N), Verb (V), Prepositions (P), Determiners (D), Adjectives (A), Complementizers (C)

- *Head*: Each phrase have a name- Noun Phrase (NP), Verb Phrase (VP), Prepositional Phrase (PP), Adjective Phrase (AP) etc. There is a main constituent in a phrase which gives it a name and thereby a Syntactic Category (NP, VP etc). This main constituent is called the Head of a phrase. Therefore, the head of a Noun Phrase is a Noun. The head of a Verb Phrase is a Verb, and so on.

- *Specifiers*: A specifier specifies or modifies the head of a phrase. For example: all, some, every, the, Boy's, his Daddy's etc. are specifiers of a Noun Phrases (NP). Very, extremely, rather, quite, etc.(also known as degree particles) are the specifiers of Adjective Phrases (AP).

- *Projection*: The nodes that dominates the head of a phrase are called the Projection of its head. For example: A VP is the projection of its head Verb (V).

- *Maximum Projection*: The higher most projection of a head is also called its maximum projection. VP is the maximum projection of V, NP is the maximum projection of N etc.

- *Intermediate Projection*: The nodes in between head and its maximum projection are called Intermediate Projections.

- *Complements*: They are the lexical categories that the head selects or chooses to combine with, to form a phrase. Complements are also known as direct objects in English. Choosing the complements depend on the heads. In some cases the head needs a complement, but sometimes not. In the sentence "He sleeps." the verb **sleeps** do not need a complement. But "He loves her." would be an incorrect sentence if **loves** is not followed by the complement **her** .

- *Adjuncts*: The optional lexical categories succeeding the complements are called Adjuncts. They are considered as optional because, even if they are left out, the sentence will be grammatically correct. Adjuncts are in a sense providers of extra informations about the action mentioned by the verb.

- *Subcategorization*: The process of head selecting it's complements and adjuncts is called Subcategorization.

Figure 1 demonstrations how the constituents which we discussed above fit in the tree representation of phrase structure.

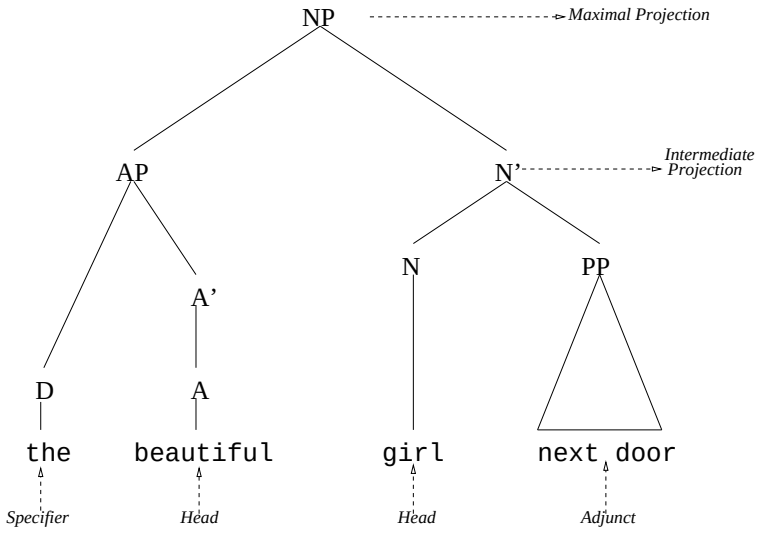


Figure 1:

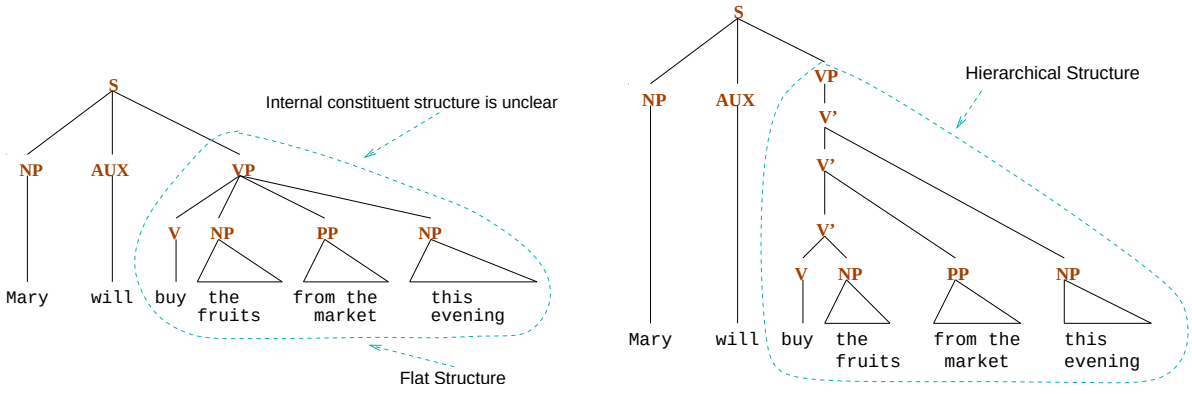


Figure 2:

3 Patterns in Phrase Structure Rules

Take a look at Figure 2(a) . The structure of the VP here is a flat one where the complements and adjuncts hang under one single node VP. It is difficult to decide which one is the complement and adjunct. This structure is problematic when we have a sentence like *"Mary will buy the fruits from the market this evening and her husband will do so tomorrow."*. Here, *"do so"* refers back to *"buy the fruits from the market"* in the sentence. It refers back to the V, NP, PP block of the sentence (excluding the last NP). But, by no means in the Figure 2(a) type of representation can we identify these three constituents as a single block. Instead, we can only refer to the whole VP as such including the last NP also. This is a problem. We need another kind of representation which solves this structural ambiguity.

Comparatively, in Figure 2(b) there is more clarity in the internal structure of the VP. We can solve our problem now by referring to the second V'(pronounced as V-bar) below the VP. That is *"do so"* will now refer back only to that V' which dominates the block containing constituents which we want . Namely, V, NP, PP and not the whole VP (that is V, NP, PP,

NP). The V'-s which we have introduced now, solves the ambiguity. Accommodating a Specifier to the head is described in Figure 3 below.

3.1 Proposing a general structure for VPs in English

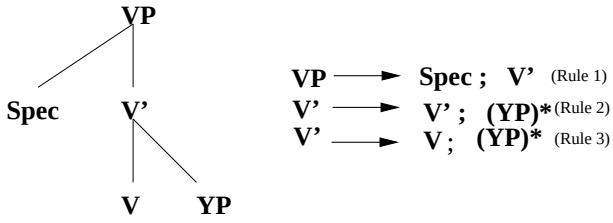


Figure 3:

In Figure 3, YP stands for Adjuncts or Complements, * stands for one or more number odYPs and brackets shows that the whole constituent can be absent as well. Note that all the branches are unary or binary. For accommodation of each adjunct, Rule 2 is repeated. In case of absence of an adjunct, Rule 2 can be omitted. With these three rules, we can generalize a common structure for PP's NP's and AP's. And hence a general pattern for all the Phrase Structures.

4 X-bar Theory for Phrase Structures

According to X-bar theory, instead of specifying separate phrase structure rules for NP's, VP's etc, the following three rules cover all the lexical categories.

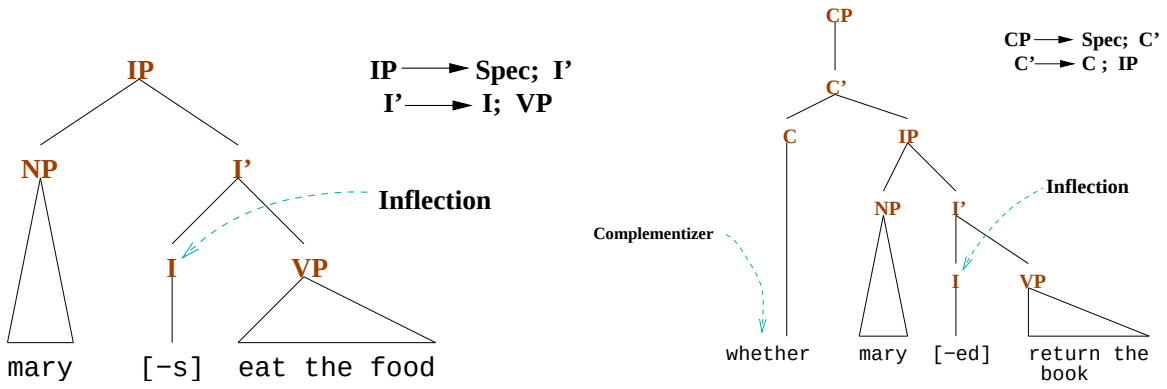
$$\text{XP} \rightarrow \text{Spec; X'}$$
$$\text{X'} \rightarrow \text{X'; (YP)*}$$
$$\text{X'} \rightarrow \text{X; (YP)*}$$

Here, XP stands for any Phrase (NP, VP, PP, AP etc.), X' stands for the intermediate projections (also called as intermediate bar-levels) and X stands for the head (N, V, P, A) Note that conjunctions (and, or) between two phrases are handled as XP $\rightarrow$ XP **Conj** XP (Eg: bread and butter). The major claims of X-bar theory [1] are : 1. All the phrase structures can be formed using the three rules mentioned above. 2. The head subcategorizes for all and only its sisters which are phrases

4.1 X-bar Scheme in Sentences and Clauses

Eg: I wonder **whether** *Mary will eat the food*

The above sentence, *"Mary will eat the food "* is another complete sentence (category **S**), which is embedded in another sentence. *"whether"* is called the **Complementizer** of *"Mary will eat the food "*. This part of the sentence (that is, *"whether Mary will eat the food".*) is a clause. In standard generative grammar, **S** together with Complementizer is labeled as **S'** with the rule **S' $\rightarrow$ Comp S'**. But the traditional rule for S as we know is **S $\rightarrow$ NP VP**. In the second rule, the head on the right hand side of the rule is undefined because, the NP



(a) (b)

Figure 4:

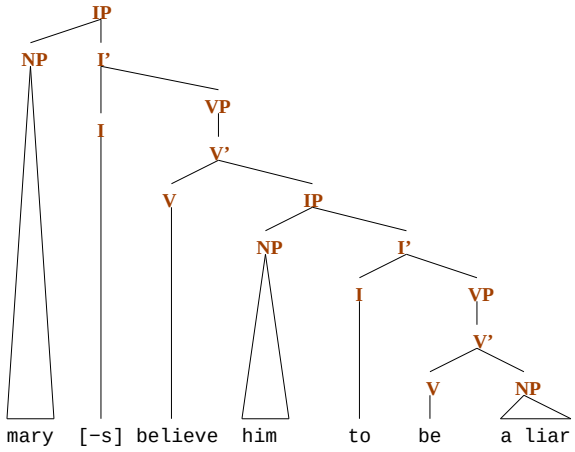


Figure 5:

and VP are phrases themselves. In the first rule, head of the clause is the sentence itself and the specifier is the complementizer. But according to X-bar theory, the head should be a word and not a sentence or a phrase. *whether* cannot be a specifier because, it is subcategorized by the preceeding verb *wonder*. At any rate head *wonder* subcategorizes only for its complements, not for specifiers. Another option is to consider *"whether Mary will eat the food"* as a Clausal Phrase (CP), with head C as *whether*. Now it fits within the frame work of X-bar scheme. But the problem remains that the head of complement of CP is a Sentence, not a word. We need an additional notion or mechanism to get rid of sentence itself being the head of a sentence. The idea is to make **Inflection (I)** which charectorizes the tense, agreement, auxiliary verb in the sentence as the head of a sentence. This makes a sentence an **Inflectional Phrase (IP)**. The sentences will then get the category IP and not S. Figures 4(a), 4(b), 5 demonstrates how IP and I works. The inflectional head [-s] results from the agreement that the of the verb with the subject (Mary which is a singular) and [-ed] results from the past tense of the verb. In case, there are no inflectational agents, the inflectional head will remain empty [_].

5 X-bar Scheme in Different Languages

English is an **S-V-O** word order language. That is, the word in sentences will be of the order: subject followed by a verb and then an object. There are languages with other word orders. For example of word orders **S-O-V**, **V-O-S**, **O-V-S** [1]. We will see how X-bar scheme works in those languages.

- For **S-V-O** (subject, verb, object) Language: (Eg: English, German)

  XP → Specifier; X'
  X' → X; Complement(s)

- For **S-O-V** languages: (Eg: Abaza - A northwest Caucasian language)

  XP → Specifier; X'
  X' → Complement(s); X

- For **V-O-S** languages: (Eg: Tzotzil - A Mayan language)

  XP → X'; Specifier
  X' → X; Complement(s)

- For **O-V-S** languages: (Eg: Hixkaryana - A Southern Guina Karib language)

  XP → X'; Specifier
  X' → Complement(s); X

6 X-bar Theory in Government & Binding Theory

According to GB, almost all the human languages have similar structure that is a part of a Universal Grammar. In GB, similar to TG, the body of grammar is divided into two blocks. D-structure (the underlying syntactic structure of the lexical entries) and S-structure (surface level representation of the sentence). GB contains further components (sub theories and rules) which constrains and conditions each block. Figure 6 explains where the X-bar theory fits in the GB theory.

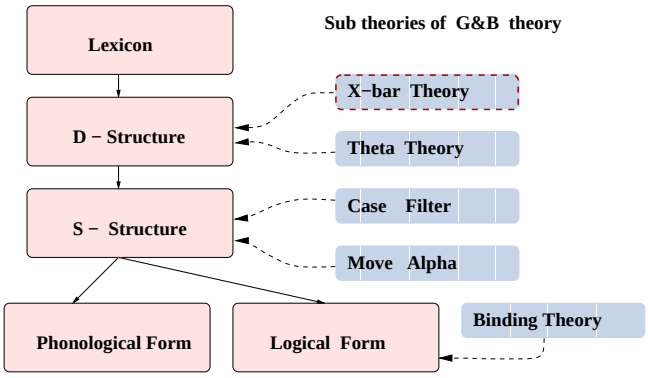


Figure 6:

7 Implementations of X-bar Theory

There are some interesting implementations of parsers based on GB theory frame work which uses X-bar scheme in its phrase structure rules.

- An interactive web based parser available online **http://www.latl.unige.ch/english/projets/ips_f.html** works on the basis of GB theory. It accepts input as sentences and outputs the syntactic structure of the sentence. This parser supports English , German, Spanish and Italian.
- PRINCIPAR: An efficient broad-coverage, principle based parser. **ftp://ftp.cs.umanitoba.ca/pub/lindek/papers/coling94.ps.gz**
- A minimal GB Parser: **http://www.cs.bu.edu/techreports/pdf/1993-013-gb-parser.pdf**

References

[1] Cheryl A. Black. A step-by-step introduction to the government and binding theory of syntax. *Summer Institute Of Linguistics. (http://www.sil.org/americas/mexico/ling/E002-IntroGB.pdf)*, 1999.

[2] Noam Chomsky. *Lectures on Government and Binding. 7th edition. Studies in Generative Grammar 9.* Berlin: Mouton de Gruyter, 1993.

[3] Liliane Haegeman. *Introduction to Government and Binding Theory. 2nd edition.* Oxford: Blackwell, 1994.

[4] Ray Jackendoff. *X-Bar Syntax: A Study of Phrase Structure, Linguistic Inquiry Monograph 2.* Cambridge, MIT Press, 1977.

[5] Peter Sells. *Lectures on Contemporary Syntactic Theories.* Stanford: CSLI Publications, 1986.