Syntax - The Study of Sentence Structure

People use language to express any idea they can think of:

"The quick brown fox jumped over the lazy dog."
"That time of year thou seest in me."
"The square of the hypotenuse is equal to the sum of the squares of the other two sides."

Human languages have two basic tools for expressing ideas – as **words** or **sentences**. Languages with a high degree of **synthesis** rely upon words, e.g. Siberian Yupik

angya-ghlla-ng-yuq-tuq
boat-augmentative-acquire-desiderative-3sing
'He wants to acquire a big boat'

Analytic languages combine words into sentences, e.g. Vietnamese

khi tôi đên nhà   bạn tôi, chúng   tôi băt dâù làm bài
when I come house friend I, PLURAL I begin do lesson
'When I came to my friend's house, we began to do lessons.'

Syntax analyzes the ways that languages combine words into larger structures such as phrases and sentences. The most fundamental observation about our word combinations is that we produce new combinations every day of our lives. We do not just repeat the same sentences over and over again. Humans around the world have the capacity to create new utterances that can be understood by any speaker of their language. Linguists would like to understand the features of language that allow this degree of creativity.

With such power comes great responsibility. We do not produce a random string of words. One fundamental question in syntax is what limits constrain the ways in which we combine words into sentences. Speakers of English recognize the difference between **acceptable** and **unacceptable** sentences:

Acceptable: Take me out to the ball game.
Unacceptable: * Take me out game the to ball.

Linguists assume that speakers of a language have internalized a grammar for their language that specifies the acceptable combinations of words. These acceptable combinations are **grammatical** because they obey the constraints described in the grammar. Anyone who has tried to learn a foreign language recognizes how difficult it is to internalize the rules of a grammar. This is why children's ability to acquire grammar unconsciously is so mysterious.

Length

We can easily expand sentences to any length. English allows:

Indefinite conjunction: "We saw a lion and a tiger and an elephant and a moth ..."
Indefinite prepositions: "the height of the lettering on the top of the page ..."
Indefinite clauses: "Mary thought that Bill said that Wendy asked whether Harry ..."

All human languages use sentences as minimal units of propositional expression, but the forms sentences can take in any language are infinitely varied. The study of sentence structure exposes the way in which human creativity is constrained by structure.

Compare the possibilities for creation in the lexicon and with sentences:

What is the longest word you know? / The longest sentence?

Can you make up a new word? / A new sentence?

Both questions reveal a level of creativity available at the sentence level, but not at the lexical level of a language. Syntax is the linguist's attempt to understand this creativity and its limits.

We know there are limits on the sentences we produce just as there are limits on the forms of words in any language. Any speaker of English would agree that "* Take me out game the ball" is not an acceptable sentence. (Linguists use asterisks to mark unacceptable sentences.) How can we produce an infinite number of original sentences on one hand, and recognize that an infinite number of sentences are unacceptable on the other?

Chomsky (1965) observed that the creativity we observe in sentence construction raises an interesting problem for language acquisition. Do children learn how to create sentences?

One hypothesis would be that children learn a language by simply imitating the sentences that other speakers produce. Imitation does not explain children's ability to produce new sentences, or accept sentences that they have never heard other speakers produce.

A second hypothesis would be that children produce new sentences by combining words together in novel ways. Random generation would predict that children would also produce lots of unacceptable sentences before they learn the correct structure for sentences.

A third hypothesis is that children produce new sentences based on analogy with the sentences they hear other speakers produce. Analogy predicts children would also produce false analogies, e.g., *Are the girls who ____ good are getting cake?   from   Are the girls getting cake?

The solution to the acquisition problem lies in recognizing sentence structure and role it plays in organizing words.

Sentence Structure

An account of sentence structure has to explain two basic features of sentences:
1. The linear order of the words in the sentence.
    A cat is on the mat   ≠   The mat is on the cat.

2. The grouping of words into **constituents**.

Constituent structure is another fundamental aspect of our grammatical knowledge. Not only can we recognize the difference between acceptable and unacceptable sentences, we can parse any acceptable sentence into smaller phrases or constituents, e.g.,

    [Any speaker of English]   [can split this sentence into two main constituents]

We can take this process further:

    [Any] [speaker of English]   [can]   [split this sentence into two main constituents]

And even further if we want to.

Linguists use **constituency tests** as evidence to back up their intuitions about sentence components. Ideally the different constituency tests produce the same results, but they can also produce different results. It takes some practice with the tests to learn how to use them properly and interpret their results.

1. **Semantic Intuition Test**—decide if the words can be grouped together semantically.

Your intuitions about the semantic relations between the words in a sentence provide a basic insight into a sentence's constituent structure.
    The phrase 'into two main constituents' specifies a result and is therefore a candidate constituent.

2. **Substitution Test**—see if a pro-form can be substituted for the proposed constituent.

Pronouns are an example of a pro-form, a form that can be substituted for a constituent of a specific type:

pronouns	substitute for	determiner phrases
do	substitutes for	verb phrases
one	substitutes for	noun phrases
there	substitutes for	prepositional phrases
auxiliaries	substitute for	auxiliary phrases

3. **Stand Alone Test**—see if the proposed constituent can stand by itself.

It helps to imagine using the constituent by itself in response to a question:
    Who can split this sentence into two main constituents?
    Any speaker of English.

    How can any speaker of English split this sentence?
    Into two main constituents.

4. **Movement Test**—see if the proposed constituent can be moved.

The movement test has more limitations than the other tests and is correspondingly more difficult to apply. One trick is to use a **pseudocleft** construction:
    [Two main constituents] is what [any speaker of English can split this sentence into].

                    as compared with

    ? Two main constituents any speaker of English can split this sentence into.

You should learn how to use the constituency tests to show that a set of words is **NOT** a constituent as well as to show that it is, e.g.,

1. Semantic Intuition
    'this sentence into two main constituents' does not make sense.

2. Substitution
    Any speaker of English can split **this** ≠
        Any speaker of English can split this sentence into two main constituents

3. Stand Alone
    What can any speaker of English split?
        *This sentence into two main constituents

4. Movement
    *This sentence into two main constituents is how any speaker of English can split.

Constituent Properties

Once we agree on the existence of sentence constituents, we can explore how they are used to construct sentences.

We can identify different types of constituents, or **phrases**, e.g.,

**Noun phrases**: *any speaker of English, this sentence, two main constituents*
Verb phrases: *split this sentence into two main constituents*
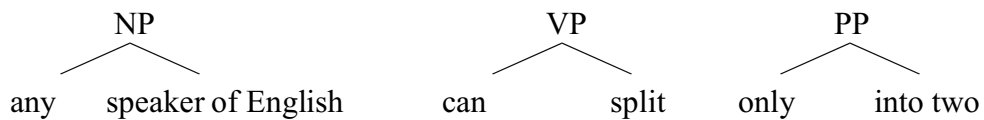**Prepositional phrases**: *into two main constituents*

The different types of phrases have some basic similarities.

One word, the **head**, is central to every type of phrase.

The head can take a **complement** phrase that provides an argument for the head, e.g.,

```
        NP                      VP                         PP
       /  \                    /  \                       /  \
  speaker  of English     split   this sentence      into   two main constituents
```
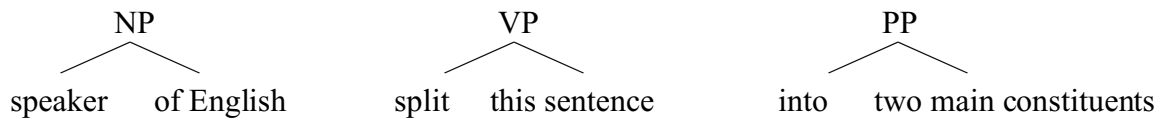
Each phrase can take a **specifier** that makes the referent more explicit, e.g.,

```
        NP                        VP                       PP
       /  \                      /  \                     /  \
    any    speaker of English  can    split         only    into two
```

Abstracting away from the specifics, we can construct a template for ALL phrases:

```
              Phrase
              /    \
                Intermediate
                 /     \
   Specifier   Head    Complement
```
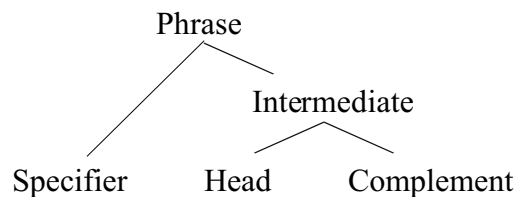
Linguists commonly borrow terms from algebra to refer to the different parts of phrase structure:

```
           XP
          /  \
             X'
            /  \
  Specifier  X   Complement
```
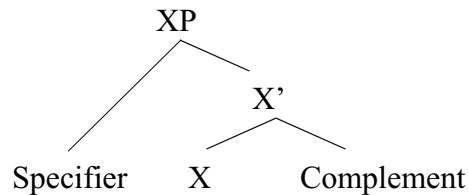
Just substitute any lexical type for X and you have an X phrase or XP!

An important aspect the so-called X-bar phrase structure is that both the specifier and complement are phrases in their own right. This feature produces the property of **recursion** that allows a rule to be repeated indefinitely. For example, noun phrases have prepositional phrase complements, and prepositional phrases take noun phrase complements, which creates structures like:

```
              NP
         ____/  \____
       Spec         N'
               ____/  \____
              N           PP
                      ___/  \___
                     P         NP
                           ___/  \___
                          N         PP
                                    |
                                   etc.
```
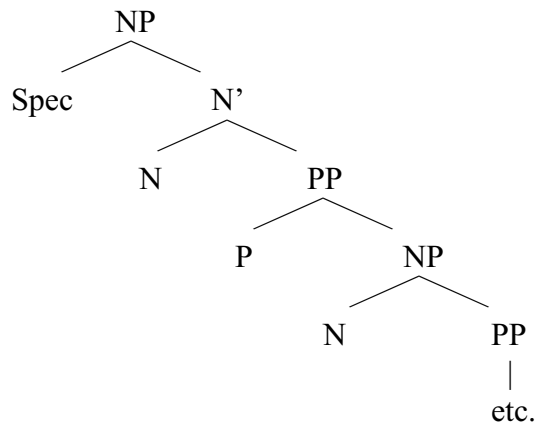
e.g., the height of the lettering on the top of the page

Recursion also allows us to add an indefinite number of adjectives to modify a noun:

```
              NP
         ____/  \____
       Spec         N'
               ____/  \____
             AdjP          N'
            /   \      ___/  \___
         Spec   Adj  AdjP        N'
                    /   \      __/  \__
                 Spec   Adj  AdjP      N
                            /   \
                         Spec   Adj
```
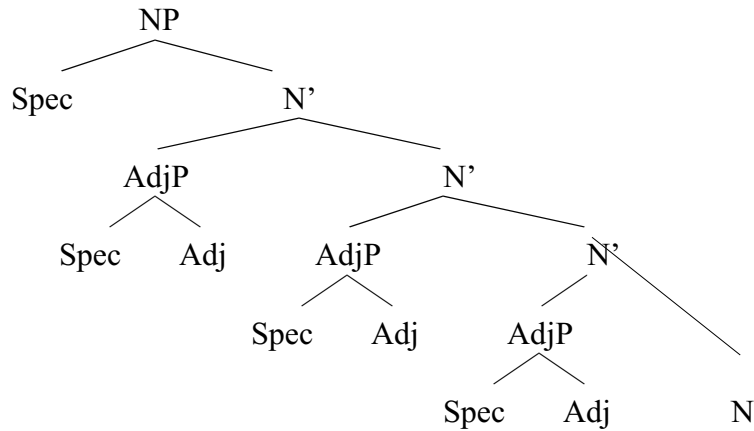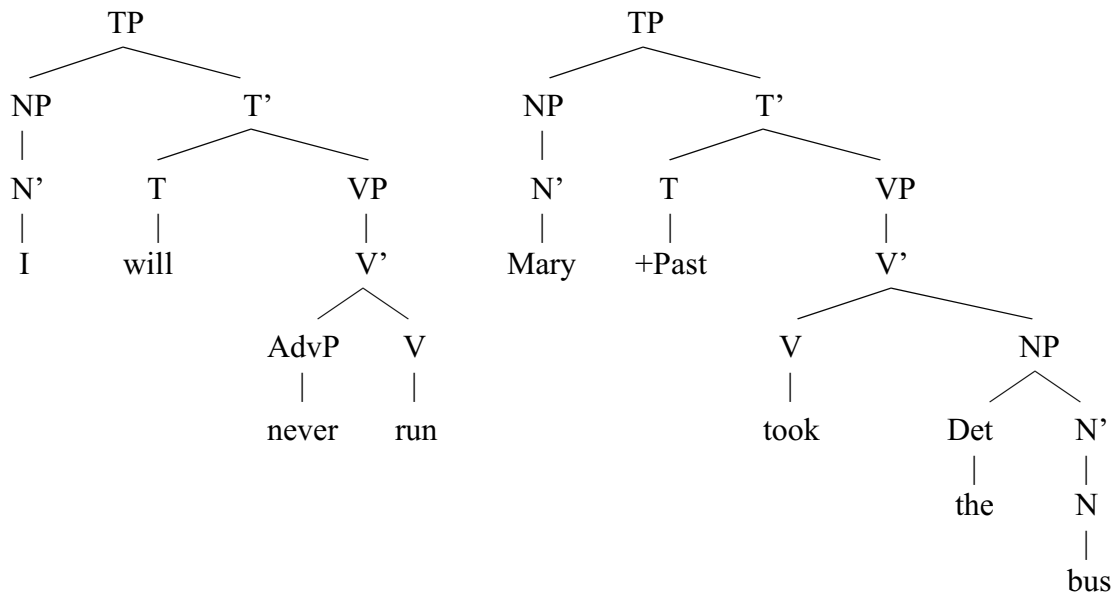
e.g., a very large, very black, rather undignified bulldog

Examples of English syntactic structures

```
         TP                          TP
      __/  \__                    __/  \__
    NP       T'                 NP       T'
    |     __/  \__              |     __/  \__
    N'   T       VP             N'   T       VP
    |    |       |              |    |       |
    I    will    V'            Mary  +Past   V'
              __/  \__                    __/  \__
            AdvP    V                    V        NP
             |      |                    |      _/  \_
           never   run                 took   Det    N'
                                               |      |
                                              the     N
                                                      |
                                                     bus
```

```
            TP                                          TP
          /    \                                      /    \
        NP      T'                                  NP      T'
       /  \    /  \                                /  \    /  \
     Det   N' T    VP                            Det   N' T    VP
      |    |  |     |                             |    |  |     |
     the   N +Past  V'                           the   N +Past  V'
           |       /  \                                |       /   \
          boy     V    NP                             boy     V'    PP
                  |   /   \                                  /  \    |
                 saw Det   N'                               V    NP  P'
                     |    /  \                              |   /  \  / \
                    the  N    PP                           saw Det N' P  NP
                         |    |                                 |  |  |   |
                        man   P'                               the N with N'
                             /  \                                  |       |
                            P    NP                               man   binoculars
                            |     |
                          with    N'
                                  |
                                  N
                                  |
                              binoculars
```
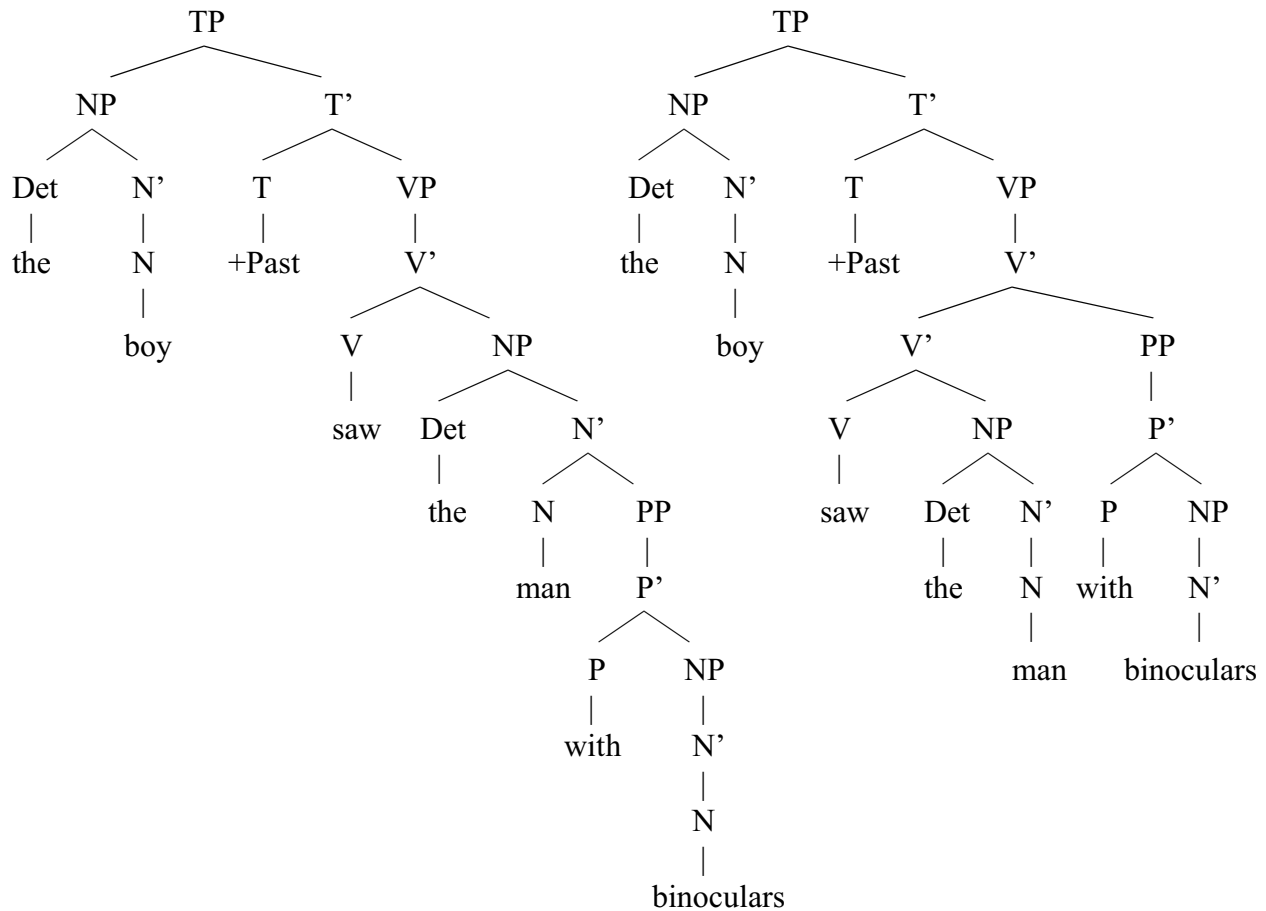
Phrase Structure Grammar

I have been using **tree structures** to display the syntactic relations between constituents. Linguists also use **phrase structure rules** to describe the same relations. A phrase structure rule looks like the phonological rules we wrote earlier. It contains a symbol on the left side of an arrow that indicates the upper or **mother node** in the tree structure, and one or more symbols on the right side of the arrow that indicate the lower or **daughter nodes** in the tree structure.

Phrase structure rules and tree diagrams are equivalent notational devices. You can convert one into the other.

We can start by writing phrase structure rules that generate our first example sentence:

   Any speaker of English can split this sentence into two main constituents.

The sentence has two main constituents, a subject and predicate, so we will need a phrase structure rule like:

   TP   NP T'

We can use X-bar theory to help us devise a phrase structure rule for the subject. We will need a specifier and a complement phrase:

NP   Det N'
N'   N PP

Note that I am assuming the prepositional phrase serves as a complement to the head of the noun phrase. The other constituent phrases can be divided into heads, specifiers and complements as follows:

TP   Tense VP
VP   V NP PP
PP   P NP

These rules constitute a **phrase structure grammar** for (a fragment of) English. The rules describe the constituent structure of our sentence, and also provide a recipe for **generating** that sentence and many more. In theory, we should be able to add new rules to our grammar until we obtained a **complete** description of the language. Such a grammar would generate every possible sentence in the language, and would not generate any unacceptable sentences. Despite 50 years of serious grammar writing, linguists still haven't produced such a grammar. We do not yet know if this problem can be solved.

You may have noticed some problems with the phrase structure rules I wrote to generate our first sentence. As written, the rules do not generate the last noun phrase 'two main constituents'. How would we need to change the rule for noun phrases to generate this phrase?

Another problem with the rules is that they do not allow for the option of not generating some constituent. Three of the noun phrases in our sentence have specifiers, but one does not. One of the noun phrases has a complement, but three do not. Linguists use parentheses in phrase structure rules to capture this feature of **optionality**. We can rewrite the noun phrase rules to allow for optional specifiers and complements.

NP   (Det) N'
N'   N (PP)

The option of adding a modifier to the noun phrase raises the need for another type of option if we add a rule for adjectival modifiers as

N'   AdjP N'

Linguists uses curly braces in phrase structure rules to capture this option.

N' $\begin{Bmatrix} \text{N (PP)} \\ \text{AdjP N'} \end{Bmatrix}$

We face two additional problems that X-bar theory creates. The phrase structure rules to generate an X-bar phrase would be:

XP   (Spec) X'
X'   X (Comp)

Our new set of rules for noun phrases provide a good example of the X-bar template. Our rule for the verb phrase is another story. Our verb phrase deviates from the X-bar template in that the

verb has two complements rather than the single complement predicted by X-bar theory. We face a direct challenge to the theory and have to decide whether we want to keep the theory and find some way to shoehorn the verb phrase into the X-bar template, or change the X-bar template to allow multiple complements.

The second problem X-bar theory creates is seen in our first phrase structure rule.

    TP   NP T'

How does this rule violate X-bar theory, and how can we change the rule to avoid such a violation?

Lexical Categories

Our phrase structure grammar still lacks one final rule for generating our sample sentence. We need a way to translate the terminal strings of our rules into actual words. The solution is simple enough; assume another rule translates the terminal strings into words of the proper lexical category.

    N   speaker, sentence, constituents
    V   split
    Aux   can
    Det   any, this, two
    P   of, into
    Adj   main

We can combine all of these rules into a single rule

    X lexeme_x

Here, lexeme points to a word in the lexicon that belongs to the lexical category X. This rule captures the idea that each word belongs to a part of speech that selects the appropriate specifier and complement phrases.

We can use several types of evidence to detect lexical categories in a language.

Notional categories

We can use word meaning to divide words into various notional categories.

    Noun - the name of a person, place or thing
    Verb - the name of an action or state
    Adjective - the name of a quality
    Preposition - the name of a path or location

The problem with such notional categories is that they are rather imprecise. For example is an action a thing or an action?

Inflectional categories

We can use lexical inflection as evidence for separating words into different lexical categories.

> Noun - inflect for plural, possession
> Verb - inflect for tense, aspect, voice, mood, agreement
> Adjective - inflect for comparative, superlative

The problem with inflectional evidence is that it is not available for every category, and some words lack an overt inflection.

Syntactic categories

We can also use syntactic frames as evidence for grouping words into lexical categories. Words are used in specific syntactic contexts, which aid in analyzing the lexical categories.

> Noun   Det _____, Det Adj _____
> Verb   Aux _____, *Please* _____
> Adjective   _____ N, Adverb _____
> Preposition   _____ NP, *right* _____
> Auxiliary   _____ V, _____ *not*

Other Languages

Our lexical category tests do not apply to other languages. We cannot use inflectional evidence in languages that lack inflection–analytic languages. Likewise, we cannot use the presence of determiners as evidence for a noun category if the language does not require determiners in its noun phrases. We can divide words in any language into notional categories, but then we have no basis for constructing syntactic rules from notional categories.

We can determine what lexical categories exist in other languages and then try to decide the extent to which we find the 'same' lexical categories in different languages.

One would think that all languages would at least contain the lexical categories of noun and verb. The Salish languages of the Northwest challenge this assumption.

Kinkade (1983, Lingua, 'Salish evidence against the universality of 'noun' and 'verb'')

Upper Chehalis only has two lexical classes: predicates and particles

> s-qʼʷət'-w-n
> continuative-burn-intrans-3sing   'fire' or 'it is burning'
>
> s-χəp-w-n
> continuative-dry-intrans-3sing   'it is drying'
>
> s-Łʔaláš   'deer' or 'it is a deer'
>
> ʔit-qʼʷalán'-č
> completive-ear-2sing   'you are all ears'

Columbian Salish

    s-q'áʔ-xn
    continuative-wedge_in-foot   'shoe' or 'it is a shoe'

The category of adjectives also varies across languages. Quechua does not distinguish between adjectives and nouns:

    chay runa hatun (kaykan)   chay runa alkalde (kaykan)
    that man big   is   that man mayor   is
    'That man is big'   'That man is mayor'

    chay hatun runa   chay alkalde runa
    'That big man'   'That mayor man' (that man who is mayor)

K'iche' expresses some adjectival notions as verbs:

    x-in-war-ik   x-in-kos-ik
    comp-1sing-sleep-status   comp-1sing-tire-status
    'I slept'   'I was tired'

    x-in-noj-ik   x-in-kikot-ik
    comp-1sing-full-status   comp-1sing-happy-status
    'I was full'   'I was happy'

Some languages (Jakaltek) lack prepositions.

Some languages have lexical categories that are not found in English.
Mayan languages distinguish a category of positionals from other lexical types:

    k-in-tak'-eʔ-ik   k-in-paq-eʔ-ik

    incomp-1sing-stand-POSITIONAL-status   incomp-1sing-climb-POSITIONAL-status
    'I am standing'   'I am climbing'

We have the following categories in languages

    Nootka (N, V)   Jakaltek (N, V, A)
    Korean (N, V, P)   English (N, V, A, P)

Linear Order

The linear order of heads, specifiers and complement phrases also varies across languages. We can describe these differences easily by means of phrase structure rules.

Korean (**Head Last**)

      S   NP VP
      VP   (PP) (NP) V   Chun ku chayk poata
      NP   (Det) N   Chun that book see
      PP   NP P   'Chun saw that book'

Selayarese (**Head First**)

      S   VP NP
      VP   V (NP) (PP)   laʔallei doeʔ iñjo ri lamari iñjo i Baso
      NP   N (Det)   take money the in cupboard the Baso
      PP   P NP   'Baso took the money in the cupboard'

Subcategories

We often need to make finer distinctions between the words in a lexical category. Transitive verbs allow an object complement, but intransitive verbs do not. Ditransitive verbs take two complement phrases. There are also verbs that take adjective and sentence complements.

    Intransitive verbs (Vi)   fall, sleep, sneeze, jog
    Transitive verbs (Vt)   throw, cut, acknowledge, seek
    Ditransitive verbs (Vdt)   give, show, put
    Adjective complement verbs be, become
    Sentence complement verbs want, promise, say

Other parts of speech have similar restrictions on the types of complements they take.

Nouns

    None   car, electricity
    PP_{to} presentation, pledge
    PP_{about} argument, discussion

Adjectives

    None   tall, smart
    PP_{about} curious, glad
    PP_{to} apparent, obvious

In addition to complement types there are other types of subcategory distinctions. Singular count nouns require a determiner, but singular mass nouns do not.

    Count nouns   table, boy, wind, mesa
    Mass nouns   water, gold, furniture

Trying to force all words into a specific lexical subcategory creates problems. Large classes of verbs regularly alternate between specific subcategories. The verbs *break*, *tear* and *drop* as well as the verbs *eat*, *drink* and *see* can appear as both transitive and intransitive verbs. English speakers differ over whether the noun *data* is mass or count. Linguists are currently attempting to describe and account for a wide range of phenomena associated with lexical subcategories.

Transformations

In addition to describing the internal structure of sentences syntax also looks at the relations between sentences. For example, we can transform a declarative sentence in English into a simple question by moving the auxiliary verb to the beginning of the sentence.

> Any speaker of English can split this sentence into two main constituents.

> becomes

> Can any speaker of English split this sentence into two main constituents?

If we maintain that these two sentences are related in some fashion, then we can try to account for the relation. We could propose a movement rule, or **transformation,** to account for the relation. Such a rule raises a number of issues.

First Guess

> Inversion Transformation   X Aux Y   Aux X Y

One problem with this version of the rule is that it is overly general. Think of an unacceptable sentence that this rule generates.

Another problem with this rule is that it doesn't explain why we move the auxiliary to the beginning of the sentence. A subpart to this problem is that the rule does not tell where to put the auxiliary once we move it.

We can solve both problems by assuming the Auxiliary moves to a position that distinguishes the illocutionary force of a question from that of a declarative sentence. We can call the node that heads this position the Question node, which projects a Question phrase. I am assuming the process of forming a question includes creating a Question phrase which then attracts the auxiliary verb. We can make a further assumption that the auxiliary verb lands at the head of the Question phrase. Our revised inversion rule would then be:

Second Guess

> Inversion Transformation   Question X Aux Y   Question-Aux X Y

The variable X in this formulation is still too general. We can tighten the rule further by specifying that X always refers to a subject.

Third Guess

This version of the rule asserts that the first auxiliary after the first noun phrase at the beginning of the sentence is the word that moves to the Question Phrase.

Other Languages

Looking at other languages we find some interesting differences in their inversion rules.

French allows an auxiliary or main verb to move to the Question node.

    Vois tu ___ le livre?
    see you ___ the book?
    'Do you see the book?'

    As tu ___ essayé?
    have you ___ tried?
    'Have you tried?'

French only allows a verb or auxiliary to move across pronominal subjects. Movement across non-pronominal subjects is not acceptable.

    *Sait Jean?
     'Knows John?'

    *A Jean essayé?
    'Has John tried?'

French speakers resort to one of the following constructions to form such questions.

    Est-ce que Jean sait?
    'Is it that John knows?'

    Jean sait-il?
    'John knows-he?'

Spanish moves the main verb to the Question node, but not auxiliary verbs. The rule applies regardless of whether or not the subject is a pronoun.

    Partió Juan ___ ?   Partió él?
    leave John ___ ?   'Left he?'
    'Did John leave?'

    *Ha Juan ___ partido?   *Ha él partido?
    has John ___ left?   'Has he left?'
    'Has John left?'

K'iche' simply introduces a particle to mark questions.

    A xel lee a Waan?

Q left the familiar John?
'Did John leave?'

The K'iche' rule resembles what happens in English when the underlying sentence does not contain an auxiliary verb. English introduces the auxiliary *do* is such situations, giving rise to a rule known as Do-Support. The French *est-ce que* construction suggests the language is moving in the direction of K'iche'.

The English Inversion rule changed during the sixteenth and seventeenth centuries. Before this time English speakers could invert the main verb.

Speak they the truth?

Cross-linguistic comparison reveals several **parameters** are necessary to account for inversion across languages. All of these languages mark yes-no questions by marking a question feature at the left periphery of the sentence. The parametric variation for inversion includes:

1. Whether or not a particle is added to mark the question.
2. Whether or not the main verb moves.
3. Whether or not the auxiliary moves.
4. Whether or not the verb can move across non-pronominal subjects.