Syntax is the area of linguistics which studies the structure of natural languages.
The underlying assumption is that languages have *structure*: not all sequences of words over the given alphabet are valid; and when a sequence of words is valid (*grammatical*), a natural structure can be induced on it.
It is useful to think of this structure as a *tree* (although we shall see other structures later).
Given a sentence in some language, not all possible trees define the structure that native speakers of the language intuitively recognize.
Natural languages have structure

Even though I *klaw* through the valley of the shadow of death,
I will *raef* no evil
Even though I *walk* through the valley of the shadow of death,
I will *fear* no evil
Even though I *ordinary* through the valley of the shadow of death,
I will *slowly* no evil
Even though I *slowly gaze* through the valley of the shadow of death,
I will *unsurprisingly do* no evil
Even though I *walk* through the valley of the shadow of death,
I will *fear* no evil
Natural languages have structure

Natural languages are *infinite*:

- The water put out the fire
- The water put out the fire, that burned the stick
- The water put out the fire, that burned the stick, that hit the dog
- The water put out the fire, that burned the stick, that hit the dog, that chased the cat

But it is possible to characterize an infinite set with finite expressions.
Intuitively, words combine to form *phrases*: 

\[(Jacob \ (served \ (seven \ years) \ (for \ Rachel))), \ and \ (they \ seemed \ to \ him \ but \ a \ few \ days \ (because \ of \ ((the \ love) \ (he \ had \ for \ her))))\].

but not:

\[(Jacob \ served) \ seven \ (years \ for) \ Rachel, \ and \ they \ (seemed \ to) \ him \ but \ (a \ few \ days \ because) \ of \ the \ love \ he \ had \ for \ her.\]

Phrases which correspond to our native speaker intuitions are called *constituents*. 
Determining constituents

The criteria for defining constituents are sometimes fuzzy. The main criterion is equivalent distribution: if two word sequences are mutually interchangeable in every context, preserving grammaticality, then both are constituents and both have the same grammatical category.
Certain grammatical operations apply only to constituents:

- **Topicalization:**
  
  *For Rachel, Jacob served seven years*

- **Cleft:**
  
  *It was for Rachel that Jacob served seven years*

- **Interjection:**
  
  *Jacob served seven years, the Bible tells us, for Rachel*
Certain grammatical operations apply only to constituents:

- **Question formation:**
  
  *How long did Jacob serve for Rachel?*

- **Coordination:**
  
  *Jacob served seven years for Rachel, and they seemed to him but a few days*

- **Anaphors refer to constituents:**
  
  *... and for Leah, too*
Inducing structure on a grammatical string is done recursively, starting with the words. To this end, words are classified into categories according to their distribution. In many languages, words are classified into substantial and functional categories.

- **substantial**: table, dogs, walked, purple, quickly
- **functional**: the, in, or

Another classification is according to whether the category is open or close.
Types of constituents

Word categories (parts of speech):

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Noun</td>
<td>table, dogs, justice, oak</td>
</tr>
<tr>
<td>V</td>
<td>Verb</td>
<td>run, climb, love, ignore</td>
</tr>
<tr>
<td>ADJ</td>
<td>Adjective</td>
<td>green, fast, mild, imaginary</td>
</tr>
<tr>
<td>ADV</td>
<td>Adverb</td>
<td>quickly, well, alone</td>
</tr>
<tr>
<td>P</td>
<td>Preposition</td>
<td>in, to, of, after, in spite of</td>
</tr>
<tr>
<td>D</td>
<td>Determiner</td>
<td>a, the, all, some</td>
</tr>
<tr>
<td>Pron</td>
<td>Pronoun</td>
<td>I, you, she, theirs, our</td>
</tr>
<tr>
<td>PropN</td>
<td>Proper Noun</td>
<td>John, IBM, University of Haifa</td>
</tr>
</tbody>
</table>
Constituents

Phrases are projections of word categories:
Noun phrases are headed by nouns:
  table → round table → the round table → the round table in the corner
  → the round table in the corner that we sat at yesterday
Verb phrases are headed by verbs:
  climbed → climbed a tree → climbed a tree yesterday
  → recklessly climbed a tree yesterday
Adjectival phrases are headed by adjectives:
  high → rather high / higher than me / high as a tree
Phrases consist of a *head* and additional *complements* and *adjuncts*. The phrase is a *projection* of its head. Complements are required by the head, and are mandatory. Adjuncts are optional, and can be iterated.

Example:  *John drinks a cup of milk every morning*
Syntactic phenomena

- Agreement
- Subcategorization
- Case assignment
- Unbounded dependencies
- Subject/object control
- Coordination
$E_0$ is a small fragment of English consisting of very simple sentences, constructed with only intransitive and transitive (but no ditransitive) verbs, common nouns, proper names, pronouns and determiners. 

Typical sentences are:

- A sheep drinks
- Rachel herds the sheep
- Jacob loves her
Similar strings are not $E_0$- (and, hence, English-) sentences:

* Rachel feed the sheep
* Rachel feeds herds the sheep
* The shepherds feeds the sheep
* Rachel feeds
* Jacob loves she
* Jacob loves Rachel she
* Them herd the sheep
There are constraints on the combination of phrases in $E_0$:

- The subject and the predicate must *agree* on number and person: if the subject is a third person singular, so must the verb be.
- Objects complement only – and all – the *transitive* verbs.
- When a pronoun is used, it is in the *nominative* case if it is in the subject position, and in the *accusative* case if it is an object.
$E_1$ is a fragment of English, based on $E_0$, in which verbs are classified to subclasses according to the complements they “require”:

Laban gave Jacob his daughter
Jacob promised Laban to marry Leah
Laban persuaded Jacob to promise him to marry Leah

Similar strings that violate this constraint are:

* Rachel feeds Jacob the sheep
* Jacob saw to marry Leah
With the addition of infinitival complements in $E_1$, $E_2$ can capture constraints of argument control in English:

- Jacob promised Laban to work seven years
- Laban persuaded Jacob to work seven years
Another extension of $E_1$ is $E_3$, typical sentences of which are:

- The shepherd wondered whom Jacob loved.
- The shepherd wondered whom Laban thought Jacob loved.
- The shepherd wondered whom Laban thought Rachel claimed Jacob loved.

An attempt to replace the gap with an explicit noun phrase results in ungrammaticality:

* The shepherd wondered who Jacob loved Rachel.
The gap need not be in the object position:

* Jacob wondered who loved Leah
* Jacob wondered who Laban believed loved Leah

Again, an explicit noun phrase filling the gap results in ungrammaticality:

* Jacob wondered who the shepherd loved Leah
Long distance dependencies

More than one gap may be present in a sentence (and, hence, more than one filler):

*This is the well which Jacob is likely to draw water from.*

*It was Leah that Jacob worked for without loving.*

In some languages (e.g., Norwegian) there is no (principled) bound on the number of gaps that can occur in a single clause.
Long distance dependencies

There are other fragments of English in which long distance dependencies are manifested in other forms.

*Topicalization:*

Rachel, Jacob loved

Rachel, every shepherd knew Jacob loved

Another example is *interrogative sentences:*

Who did Jacob love?
Who did Laban believe Jacob loved?
Coordination is accounted for in the language fragment $E_4$:

No man lift up his [hand] or [foot] in all the land of Egypt
Jacob saw [Rachel] and [the sheep of Laban]
Jacob [went on his journey] and [came to the land of the people of the east]
Jacob [went near], and [rolled the stone from the well’s mouth], and [watered the flock of Laban his mother’s brother]. every [speckled] and [spotted] sheep
Leah was [tender eyed] but [not beautiful]
[Leah had four sons], but [Rachel was barren]
She said to Jacob, “[Give me children], or [I shall die]!”
The goals of syntactic analysis

Given a natural language sentence, syntactic analysis provides a structural description of the sentence. To do so, one must have a model of the structure of the language. Syntax is concerned with a formulation of the structure of natural languages. An example of a syntactic formalism is context-free grammars. In CFGs, the structure of sentences is modeled by derivation trees.