

1. Formale Sprachen

Die Kommunikation zwischen Menschen und Maschinen wird heutzutage meist über Programmiersprachen oder Skripte geregelt. Doch in der jüngeren Vergangenheit können wir immer mehr Gespräche mit Maschinen führen.

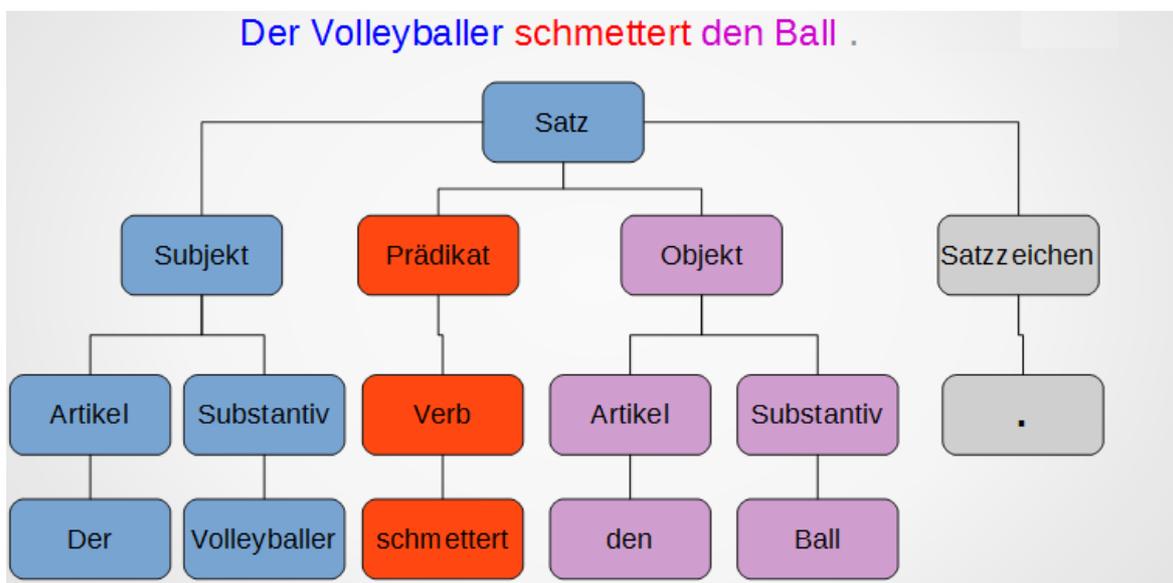
Es stellt sich dabei nun die Frage, wie bringt man einer Maschine bei eine Sprache zu verstehen. Um uns diesem Problem zu nähern schauen wir im folgenden Kapitel das Konzept der formalen Sprachen an.

1.1 Grundbegriffe

Bevor wir verstehen können, wie man Sprache einem Computer beibringt, müssen wir uns die Sprachen erst einmal genau betrachten. Wir beginnen mit einem einfachen Satz:

„Der Volleyballer schmettert den Ball.“

Zerlegt man diesen Satz in seine Bestandteile sieht das so aus:



Wir erkennen eine Baumstruktur, die wir den **Syntaxbaum** nennen. Es gibt offensichtlich zwei Arten von Knoten in diesem Baum, die unterschiedliche Bedeutungen haben:

1. Alle inneren Knoten, die „allgemeine“ Begriffe haben und bei einem anderen Satz durch ein anderes Wort ersetzt werden könnten. (**Nicht-Terminalsymbole**)
2. Alle Blätter des Baumes, die konkrete Wörter enthalten, die nicht mehr „weiterentwickelt“ werden können. (**Terminalsymbole**)

Aus dem Baum können wir ein paar Regeln zur Bildung eines Satzes ableiten.

R1: Satz → Subjekt Prädikat Objekt Satzzeichen

R2: Subjekt → Artikel Substantiv

R3: Prädikat → Verb

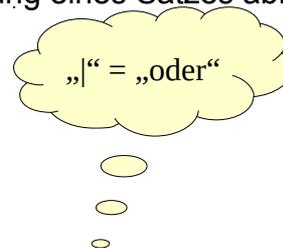
R4: Objekt → Artikel Substantiv

R5: Artikel → „Der“ | „den“

R6: Substantiv → „Volleyballer“ | „Ball“

R7: Verb → „schmettert“

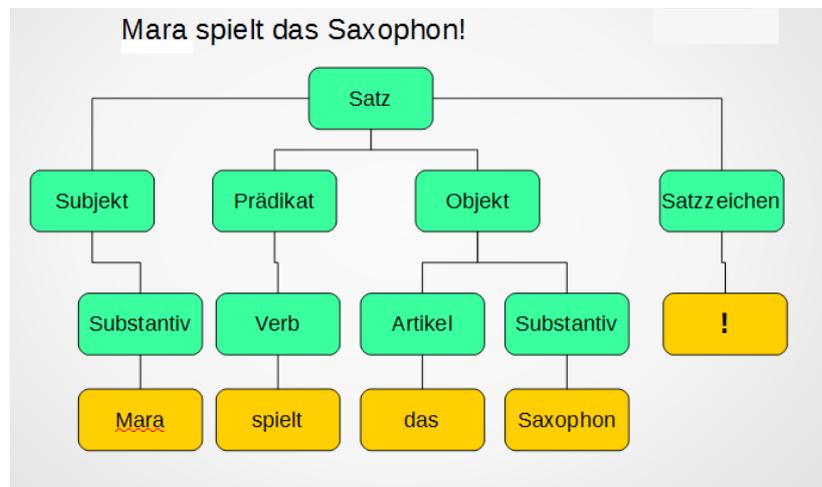
R8: Satzzeichen → „.“



Wir nennen diese Ableitungsregeln eine Grammatik, welche die Syntax der Sprache festlegt. Die Syntax überprüft allerdings nur, ob die Struktur passt. Ob der Satz Sinn hat, entscheidet die Semantik. Wenn die Syntax stimmt, muss die Semantik noch lange nicht passen, wie wir sehen werden, wenn wir einen zweiten Satz hinzufügen:

„Mara spielt das Saxophon!“

Wir sehen rechts den Syntaxbaum. In Orange sind die Terminalsymbole und in Türkis die Nicht-Terminalsymbole.



Syntaxbaum des zweiten Satzes

Abgeleitet ergibt sich eine Erweiterung auf folgende Regeln:

- R1: Satz → Subjekt Prädikat Objekt Satzzeichen
- R2: Subjekt → Artikel Substantiv | **Substantiv**
- R3: Prädikat → Verb
- R4: Objekt → Artikel Substantiv
- R5: Artikel → „Der“ | „den“ | „das“
- R6: Substantiv → „Volleyballer“ | „Ball“ | „**Saxophon**“ | „**Mara**“
- R7: Verb → „schmettert“ | „**spielt**“
- R8: Satzzeichen → „.“ | „!“

Nach dieser Grammatik ist aber auch der Satz: „den Mara schmettert Das Ball.“ Ein gültiger Satz, der semantisch jedoch sinnfrei ist.

Eine Zusammenstellung aus allen Terminalsymbolen nennen wir Alphabet. Das Alphabet zusammen mit der Grammatik und den Nicht-Terminalsymbolen ergibt eine **formale Sprache**.

Im obigen Beispiel wäre das Alphabet:

$A = \{ \text{„Der“}, \text{„den“}, \text{„das“}, \text{„Volleyballer“}, \text{„Ball“}, \text{„Saxophon“}, \text{„Mara“}, \text{„schmettert“}, \text{„spielt“}, \text{„.“}, \text{„!“} \}$

Einzelne Elemente des Alphabets nennen wir oft auch Zeichen oder Symbol und deren Aneinanderreihung eine Zeichenkette. Dies macht besonders Sinn, wenn man formale Sprachen wie z.B. Chemische Formeln (z.B. C_2H_5OH) oder Funktionsgleichungen in der Mathematik (z.B. $f(x) = x^2 - 2x + 3$) betrachtet.

Man beachte: Ein Alphabet gibt alle möglichen Bestandteile einer Sprache an, aber nicht alle Kombinationen dieser Bestandteile sind teil der formalen Sprache!

Aufgaben:

1) Beim Programmieren von Webseiten verwendet man HTML als Sprache. Dabei werden Farben durch einen Hexadezimalen Code dargestellt.

Der Code beginnt immer mit einem „#“.

Anschließend folgt ein Farbcode bei dem je zwei Stellen den Rot-, Grün- und Blauanteil darstellt.

Bsp: #9900BB

Hexadezimal bedeutet, dass die Zahl im 16er System geschrieben wird. Daher kann jede Stelle der Zahl auch ein „A“ = 10, „B“ = 11, ... „F“ = 15 sein. Die Zahl „FF“ entspricht also $16 \cdot 15 + 1 \cdot 15 = 255$.

Finden Sie alle Syntaxregeln der Farbdefinition und Zeichnen Sie einen Syntaxbaum für das Bsp. Nennen Sie alle Terminal- und Nicht-Terminalsymbole der Sprache und geben Sie das Alphabet an.