

Formale Sprachen

vorausgesetzt: ein **Alphabet** Σ , das ist eine endliche Menge von Symbolen.

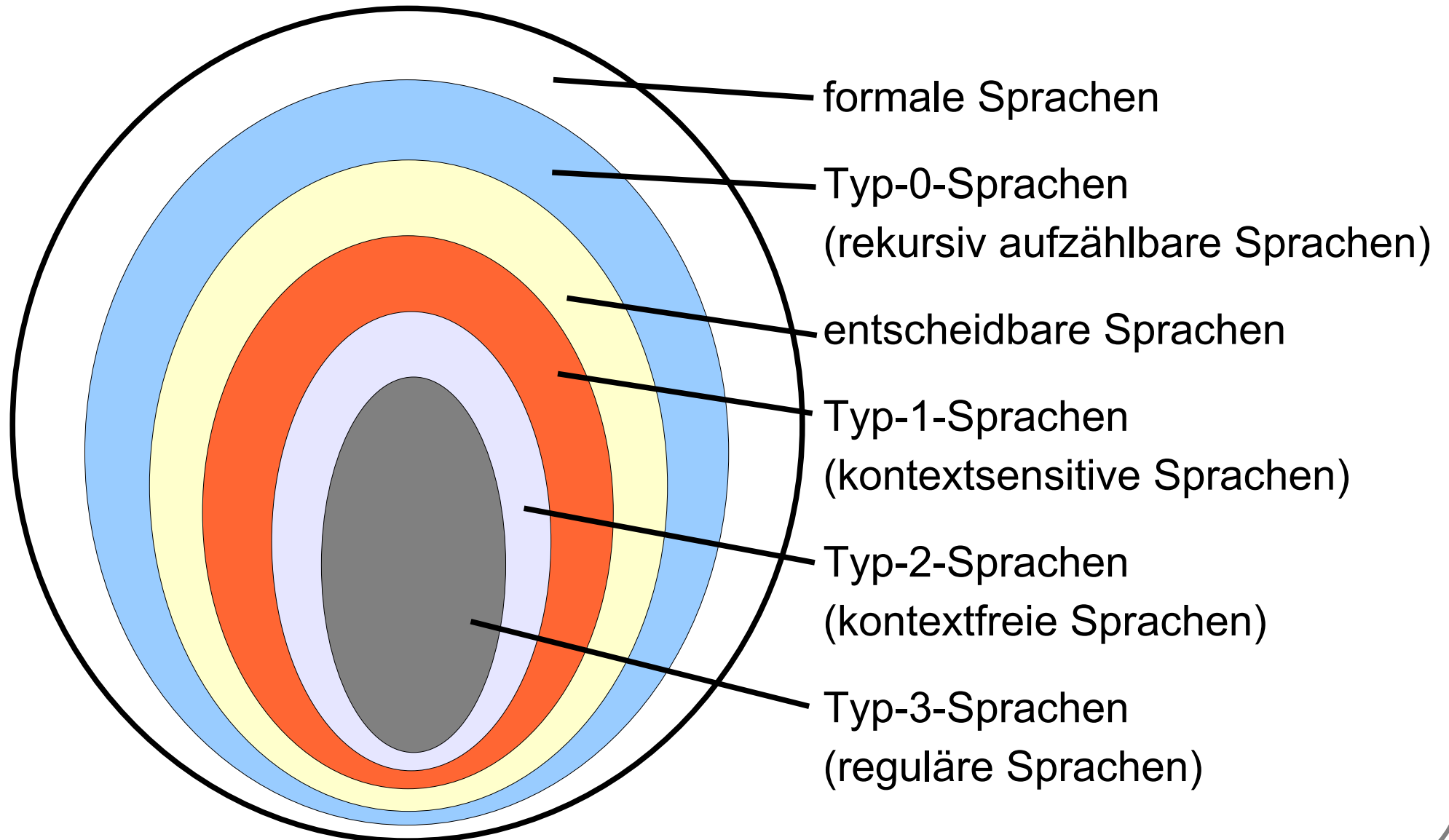
Definitionen: w ist ein **Wort** über dem Alphabet Σ genau dann, wenn w eine endliche Folge von Symbolen aus Σ ist.

Das leere Wort $\varepsilon :=$ die leere Folge \emptyset ($\varepsilon \notin \Sigma$).

Σ^* („der Kleene-Abschluss von Σ “) = die Menge aller Wörter über Σ .

L ist eine **(formale) Sprache** über dem Alphabet Σ genau dann, wenn L eine Teilmenge (\subseteq) von Σ^* ist.

Die Chomsky-Hierarchie



Erzeugungsgrammatiken

Eine **allgemeine Regelgrammatik** $G = \langle \Sigma, V, S, P \rangle$ (Typ-0-Grammatik) besteht aus

- einer endlichen Menge Σ von Symbolen (**Terminalsymbole**, z.B. Wortformen),
- einer endlichen Menge V von Variablen (**Nicht-Terminalsymbole**, z.B. Kategoriensymbole), wobei Σ und V disjunkt sind, d.h. keine gemeinsamen Elemente haben,
- einem **Startsymbol** $S \in V$,
- einer endlichen Menge P von Produktionen (**Ersetzungsregeln**) $l \rightarrow r$, wobei die linke Seite l ein beliebiges nicht-leeres Wort und die rechte Seite r ein beliebiges Wort über dem Gesamtalphabet $\Sigma \cup V$ ist ($r = \epsilon$ ist also zugelassen, $l = \epsilon$ hingegen nicht).

Ableitungen und erzeugte Sprachen

Ein Wort w' ist in $G = \langle \Sigma, V, S, P \rangle$ aus dem Wort w **ableitbar** ($w \Rightarrow^*_G w'$) genau dann, wenn es eine Folge von Produktionen gibt, die w' schrittweise aus w ableiten.

$L(G) := \{w \in \Sigma^* \mid S \Rightarrow^*_G w\}$ = die Menge der Wörter über Σ (das ist die Menge der Terminalsymbole von G), die in G aus dem Startsymbol von G ableitbar sind, heißt **die von G erzeugte Sprache**.

Ableitungsbäume (Syntaxbäume)

Wenn in G keine Mehrfachersetzungen vorkommen, wenn also in jeder Regel $l \rightarrow r$ nur jeweils ein Symbol von l ersetzt wird, dann können zu den Ableitungen **Ableitungsbäume (Syntaxbäume)** konstruiert werden.

Durch eine Ableitung wird dem resultierenden Wort – einer Folge von Terminalsymbolen – eine **(syntaktische) Struktur** aufgeprägt; und umgekehrt lässt sich jeder Ableitungsschritt als Zerlegung des Wortes in Konstituenten auffassen.

Ein Ableitungsbaum lässt sich immer ohne überkreuzende Kanten darstellen.

Beachte: Die Übersetzung einer Ableitung in einen Baumgraphen und die Rückübersetzung ist nur unter besonderen Bedingungen eindeutig!