§ 3 Formal languages

3.1 Def.: (a) Let Σ be an alphabet, i.e. a finite set of letters.

(b) A subset $\mathcal{L} \subseteq \Sigma^* := \bigcup_{n\in\mathbb{N}} \Sigma^n$ is a language.

(c) A Semi Thue system is a finite set $R \subseteq \Sigma^* \times \Sigma^*$ of rules.

For $u, v \in \Sigma^*$, we write ("how to transform words")

$$u \xrightarrow{0}_R v \;:\Longleftrightarrow\; u = v$$

$$u \xrightarrow{1}_R v \;:\Longleftrightarrow\; \exists w_1, w_2 \in \Sigma^* \; \exists (r,s) \in R : u = w_1 r w_2 \wedge v = w_1 s w_2$$

$$u \xrightarrow{n+1}_R v \;:\Longleftrightarrow\; \exists w \in \Sigma^* : u \xrightarrow{n}_R w \wedge w \xrightarrow{1}_R v$$

$$u \xrightarrow{*}_R v \;:\Longleftrightarrow\; \exists n \in \mathbb{N} : u \xrightarrow{n}_R v$$

(d) A Chomsky grammar is $\mathcal{G} = (\Sigma, \Omega, S, R)$ with

- Σ alphabet
- $\Omega \subseteq \Sigma$ finite subset, the terminal alphabet
- $S \in \Sigma$ the start symbol
- $R \subseteq \Sigma^* \times \Sigma^*$ semi Thue system

$\mathcal{L}_{\mathcal{G}} := \{ w \in \Omega^* \mid S \xrightarrow{*}_R w \}$ the language generated by $\mathcal{G}$, a Chomsky language ("type 0 language").

3.2 Lemma: Let Σ be "gödelized", i.e. for any $b \in \Sigma$ we have a Gödel number $\ulcorner b \urcorner$, and $w = b_1 \dots b_m$ is encoded as $\langle \ulcorner b_1 \urcorner, \dots, \ulcorner b_m \urcorner \rangle \in \mathbb{N}$ (see 1.11).
If $\mathcal{L}$ is a Chomsky language, then $\mathcal{L} \in \Sigma_1^0$, i.e. $\mathcal{L}$ rec. enumerable. (see Def. 2.16).

Proof: $w \in \mathcal{L} \Longleftrightarrow w \in \Omega^* \wedge S \xrightarrow{*}_R w$

$$w \in \Omega^* \Longleftrightarrow \underbrace{w \in Seq}_{PRP\,(1.12)} \wedge \underbrace{\forall i < \underbrace{length(w)}_{PRF\,(1.12)} : \underbrace{(w)_i \in \Omega}_{PRP}}_{PRP\,(1.10)} \quad \rightsquigarrow \Delta_0^0 = PRP$$

(Ω finite)

$$u \xrightarrow{*}_R v \Longleftrightarrow \exists f : \underbrace{\Big[Seq(f) \wedge (f)_0 = u \wedge (f)_{length(f) \dot- 1} = v \wedge \underbrace{\forall i < length(f) \dot- 1 : \underbrace{(f)_i \xrightarrow{1}_R (f)_{i+1}}_{PRP}}_{PRP\,(1.10)} \Big]}_{PRP} \quad \rightsquigarrow \Sigma_1^0 \; (2.23)$$

$$u \xrightarrow{1}_R v \Longleftrightarrow \underbrace{\exists w_1 < u \; \exists w_2 < u \; \exists r < u \; \exists s < v}_{PRP\,(1.10)} : \underbrace{(r,s) \in R}_{\text{finite, hence PRP}} \wedge u = w_1 r w_2 \wedge v = w_1 s w_2$$

□

3.3 Theorem: $\{ \mathcal{L} \text{ Chomsky language} \} = \Sigma^0_1$.

Proof: „$\subseteq$" 3.2

„$\supseteq$" $P \in \Sigma^0_1 \overset{2.23 \,\&\, 2.16}{\Longrightarrow} \exists f : \mathbb{N} \to \mathbb{N}^n : P = rg(f),\ f_1, \dots f_n \in \mathbb{F}$.

Construct a Turing machine based on a Chomsky grammar and prove that for any $f \in \mathbb{P} = \mathbb{P}_{Turing}$ we find such a Turing machine whose output is f. $^{2.8}$ Use $\mathbb{F} \subseteq \mathbb{P}$. □

3.4 Corollary: The general word problem for semi Thue systems is undecidable, if $|\Sigma| \geq 2$, i.e. it is $\Sigma^0_1 \setminus \Delta^0_1$.

Proof: 1.) By Thm. 3.3, there is a language $\mathcal{L} \in \Sigma^0_1 \setminus \Delta^0_1$. Then $w \in \mathcal{L}$ is a $\Sigma^0_1 \setminus \Delta^0_1$ problem.

2.) Encode the letters of Σ as words in $\mathcal{L}$. □

3.5 Remark: (a) A grammar $\mathcal{G} = (\Sigma, \Omega, S, R)$ is context sensitive (type 1), if all rules in R are of the form $(wAv, wuv) \in R$, $A \in \Sigma \setminus \Omega$, $u \in \Sigma^* \setminus \{\emptyset\}$.

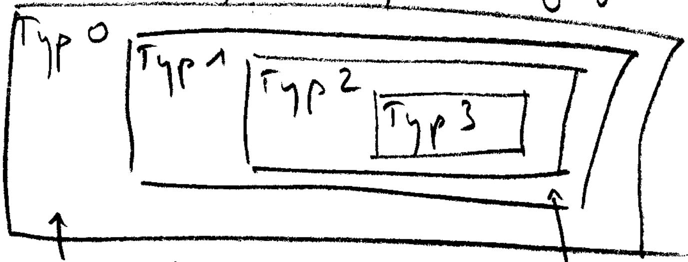
The word problem for type 1 languages is Δ^0_0, so decidable.

(b) A grammar is context free (type 2), if $(A, u) \in R$, $A \in \Sigma \setminus \Omega$, $u \in \Sigma^* \setminus \{\emptyset\}$

We can always bring it in Chomsky normal form, i.e. rules are of the form $(A, BC) \in R$, $A, B, C \in \Sigma \setminus \Omega$ or (A, σ), $\sigma \in \Omega$.

(c) A grammar is regular (type 3), if the rules are (A, σ) or $(A, \sigma B)$, $A, B \in \Sigma \setminus \Omega$, $\sigma \in \Omega$.

(d) The Chomsky hierarchy of languages is:



word problem is $\Sigma^0_1 \setminus \Delta^0_1$ (undecidable)

word problem is Δ^0_0 (decidable)