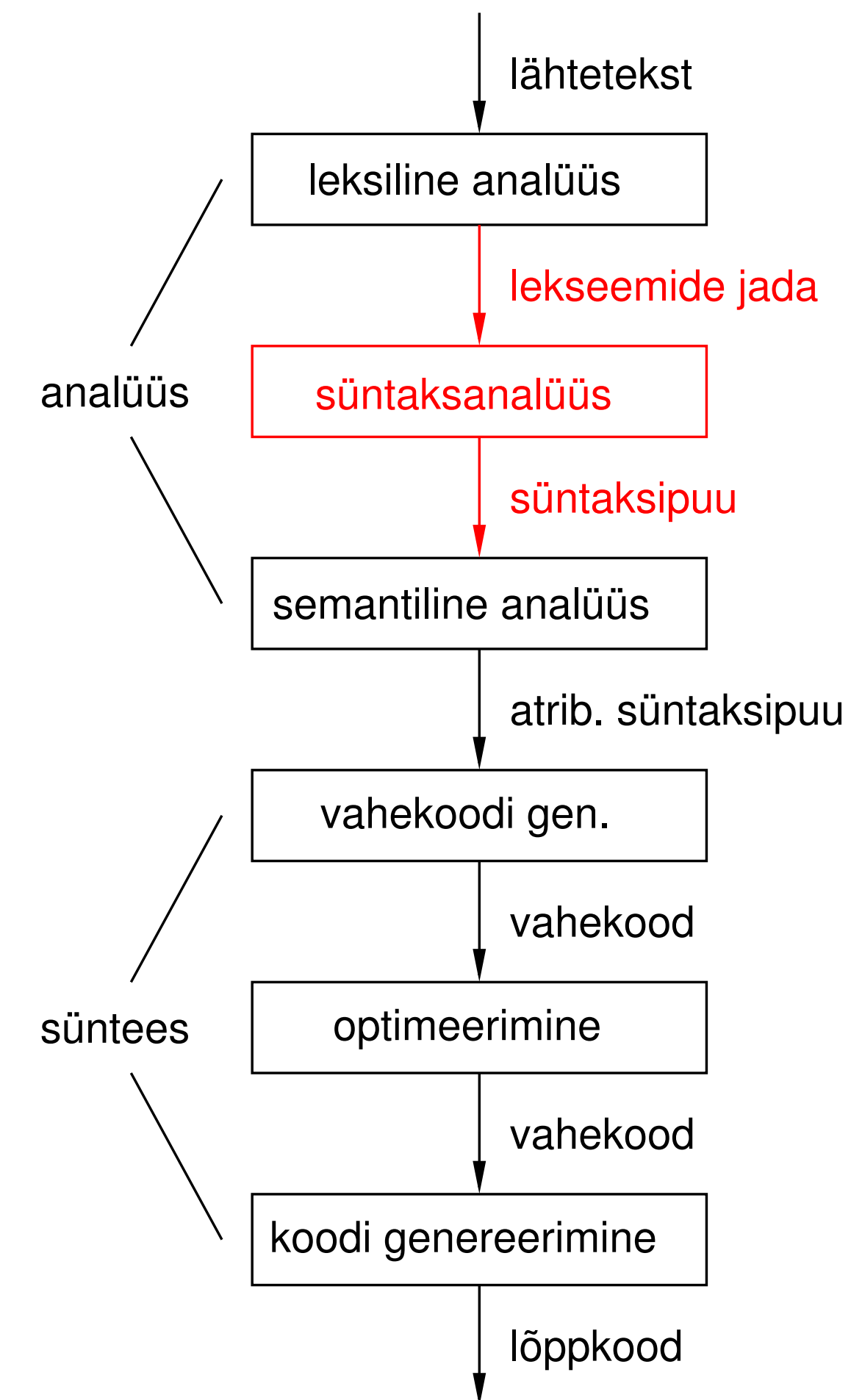


Süntaksanalüüs

Sissejuhatus
Grammatikad

Süntaksanalüüs

- **Süntaksanalüüs** kontrollib programmi struktuuri vastavust keele grammatikale:
 - saab sisendina, skanneri poolt genereeritud, lekseemide jada;
 - väljastab programmi esitava (abstraktse) süntaksipuu;
 - süntaktiliste vigade korral, teeb kindlaks nende asukoha;
 - ...teavitab võimalikest vea põhjustest;
 - ...püüab veast toibuda ja jätkata analüüsi (et järgnevaid vigu avastada).
- Süntaksanalüüsi kutsutakse **parsimiseks** (**parsing**) ning vastavat analüsaatorit nimetatakse **parseriks** (**parser**).



Grammatikad

- Keelte süntaksi kirjeldatakse reeglina kontekstivaba grammatika abil.
- **Grammatika** on nelik $G = \langle N, T, P, S \rangle$, kus
 - N on lõplik **mitteterminaalide** tähestik;
 - T on lõplik **terminaalsümbolite** tähestik;
 - $N \cap T = \emptyset$ ja $V = N \cup T$;
 - $P \subset \{\alpha \rightarrow \beta \mid \alpha \in V^+, \beta \in V^*\}$ on lõplik **produksioonireeglite** hulk;
 - $S \in N$ on **algsümbol**.
- Grammatika on **kontekstivaba** (**context-free**), kui produktsioonireeglid on kujul $A \rightarrow \alpha$, kus $A \in N$ ja $\alpha \in V^*$.

$Exp \rightarrow Exp + Exp$

$Exp \rightarrow Exp - Exp$

$Exp \rightarrow Exp * Exp$

$Exp \rightarrow Exp / Exp$

$Exp \rightarrow \mathbf{num}$

$Exp \rightarrow (Exp)$

Grammatikad

- Jada $w \in V^*$ nimetatakse **lausevormiks** (sentential form).
- Lausevorm $v \in V^*$ on **otsetuletatav** (directly derivable) lausevormist $u \in V^*$ (tähistus $u \longrightarrow v$), kui leiduvad $w_1, w_2, \alpha, \beta \in V^*$ sellised, et $u = w_1\alpha w_2$, $v = w_1\beta w_2$ ja $\alpha \rightarrow \beta \in P$.
- Relatsiooni \longrightarrow refleksiivset transitiivset sulundit (tähistus \Longrightarrow) nimetatakse **derivatsiooniks** (derivation) ehk **tuletuseks**.
- Grammatika $G = \langle N, T, P, S \rangle$ genereerib **keele**

$$L(G) = \{w \in T^* \mid S \Longrightarrow w\}$$

- Grammatikad G_1 ja G_2 on **ekvivalentsed**, kui $L(G_1) = L(G_2)$.

Grammatikad

Chomsky hierarhia:

	Produktsioonid	Keelte tüüp	Automaat
L_0	$\alpha \rightarrow \beta$	Semi-Thue süsteemid	Turingi masin
L_1	$\alpha A \beta \rightarrow \alpha \gamma \beta$	Kontekstist sõltuvad	Tõkestatud Turingi masin
L_2	$A \rightarrow \alpha$	Kontekstivabad	Magasinmäliga automaat
L_3	$A \rightarrow w, A \rightarrow wB$	Regulaarsed	Lõplik automaat
(L_4)	$A \rightarrow w$	Lõplikud	Tsükliteta lõplik automaat

kus $A, B \in N$, $\alpha, \beta, \gamma \in V^*$ ja $w \in T^*$.

Lemma: Chomsky hierarhia on range; so.:

$$(L_4) \subset L_3 \subset L_2 \subset L_1 \subset L_0$$