CS 5800 Notes

based on T. A. Sudkamp [2]

1 The Chomsky hierarchy

Table 1 gives an overview of the hierarchy of languages studied in this course, and their corresponding grammars and machines. This includes the families of grammars labeled by Chomsky as type 0, 1, 2 and 3, deriving the r.e. languages, CSLs, CFLs and regular languages, respectively (see the Chomsky hierarchy table p. 336 in Sudkamp [2].

Table 1: Chomsky hierarchy of languages, grammars and machines

Languages	Grammars	Accepting machines
r.e.	type 0 (phrase-structure, unrestricted)	TM, DTM, NTM (non-deterministic)
recursive		TM that halts for all input
CSL	type 1 (CSG, non-contracting)	LBA (Linear-bounded automaton)
CFL	type 2 (CFG)	PDA (Push-down automaton)
regular	type 3 (regular, right-linear, left-linear)	DFA, NFA, NFA-λ

Recall that a CSG (context-sensitive grammar) is defined with the type of grammar rules $u \to v$ with $|u| \leq |v|$, implying that $v \neq \lambda$. Thus according to the definition, a CSL does not contain the empty string, so that all CFLs and regular languages that do contain the empty string are not CSLs. Apart from this fact, the Chomsky hierarchy languages fall into a strict hierarchy, in the sense that the type $i+1$ family $\subsetneq$ type i, for $i = 0, 1, 2$. Consequently, examples of type i languages can be given that do not belong to type $i+1$.

The accepting machine for CSLs (context-sensitive languages) is the LBA (linear-bounded automaton), which is defined similarly to a TM language acceptor. The LBA has the same six elements as the TM tuplet, and two additional elements: a left and right endmarker, $\langle$ and $\rangle$, which are symbols in Σ and written on the tape in the input configuration ($q_0\langle w\rangle$), to delimit the region for the computation on the tape. The endmarkers designate the left and right boundaries of the tape, and cannot be erased or traversed by the computation.

Theorem 1.1. *If L is a CSL, the L is accepted by an LBA.*
If L is accepted by an LBA, then $L - \{\lambda\}$ is a CSL.

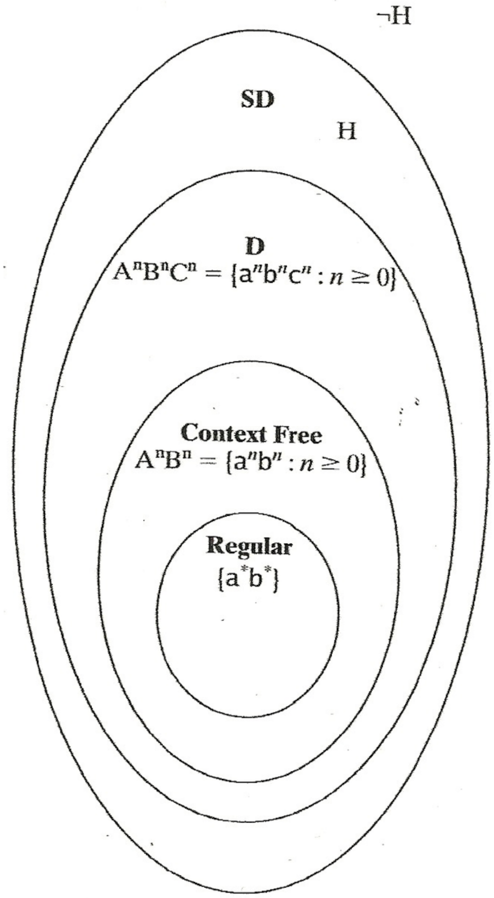

Figure 1: Hierarchy of families of languages [1]

Example 1.1. $L = \{a^n b^n c^n \mid n \geq 0\}$ is accepted by an LBA; $L - \{\lambda\} = \{a^n b^n c^n \mid n > 0\}$ is a CSL.

Since L is accepted by an LBA it is also recursive and, as a recursive language, L is r.e.

The strict hierarchy is further illustrated in Fig. 1 (from [1]), which depicts the hierarchy of the families of the r.e. languages, recursive languages, CFLs and regular languages. The family (or set) of the r.e. languages is denoted SD (semi-decidable), and the set of the recursive languages is denoted D (decidable).

The language $\{a^n b^n \mid n \geq 0\}$ denoted by $A^n B^n$ is a CFL, and is not regular (as shown with the pumping lemma for regular languages). The language $\{a^n b^n c^n \mid n \geq 0\}$ denoted by $A^n B^n C^n$ is a not a CFL, as shown with the pumping lemma for CFLs, but is recursive (in D). The halting language, H $(= L_H)$, is r.e. but not recursive (thus $H \in SD - D$) through the undecidability of the halting problem. The complement of H denoted $\neg H = \overline{H} = \overline{L_H}$ is not r.e. (see the chapter on Undecidability).

References

[1] Rich, E. *Automata, Computability, and Complexity: Theory and Applications*. Pearson Prentice Hall, 2008. ISBN: 0-13-228806-0; ISBN: 978-0-13-228806-4.

[2] Sudkamp, T. A. *An Introduction to the Theory of Computer Science − Languages and Machines*. Pearson, Addison Wesley, 3rd edition, 2006. ISBN 0-321-32221-5.