**The Compiler – Formal Languages**

* High-level language → compiler → assembly language
* Assembly language
  + Simple structure
  + Easy to parse
  + Straight-forward, unambiguous translation to machine language
* High-level language
  + Complex structure
  + Harder to recognize/parse
  + Usually no single translation to machine language
  + A formal theory of string recognition is needed to handle the complexity
* **Alphabet** – a finite set of symbols, e.g. ∑ = {a, b, c}
* **String/word** – a finite sequence of symbols (from ∑), e.g. a, ab, acb etc.
  + Length of word – # of chars in the word, e.g. |aba| = 3
  + Empty word – an empty sequence of symbols – denoted by ε, e.g.|ε| = 0