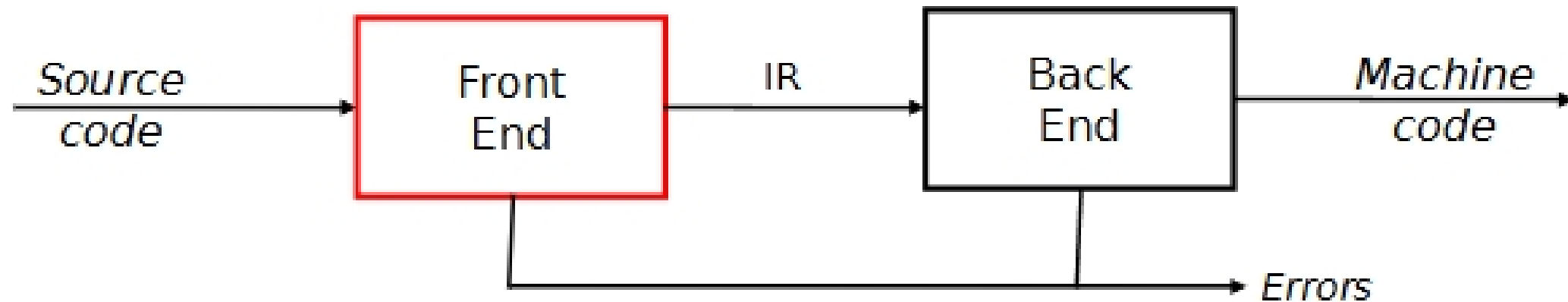




Lexical Analysis - An Introduction

The Front End

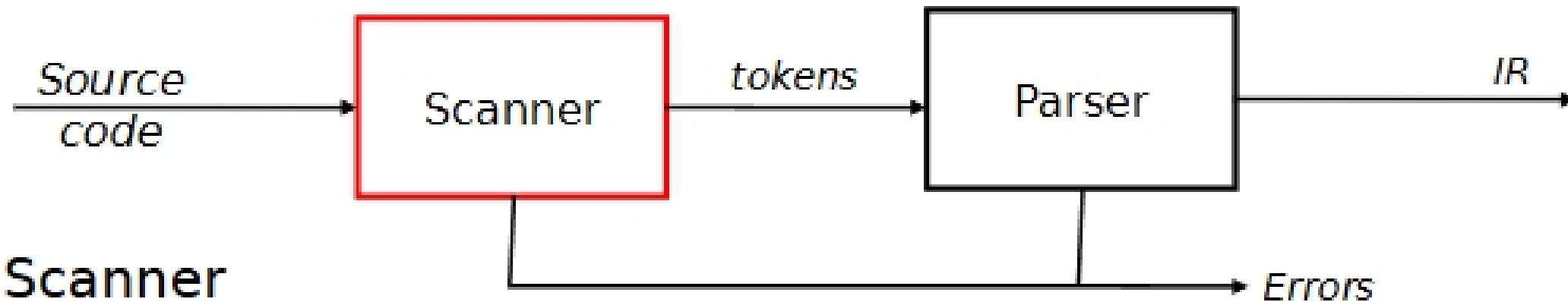


The purpose of the front end is to deal with the input language

- Perform a membership test: $\text{code} \in \text{source language}$?
- Is the program well-formed (semantically) ?
- Build an IR version of the code for the rest of the compiler

The front end is not monolithic

The Front End



Scanner

- Maps stream of characters into words

- Basic unit of syntax
- $x = x + y ;$ becomes

$\langle id, x \rangle \langle eq, = \rangle \langle id, x \rangle \langle pl, + \rangle \langle id, y \rangle \langle se, ; \rangle$

- Characters that form a word are its *lexeme*
- Its *part of speech* (or *syntactic category*) is called its *token type*
- Scanner discards white space & (often) comments

Speed is an issue in scanning

use a specialized recognizer