**Introdução - MOCC (My Own C Compiler)**

Este relatório descreve o desenvolvimento de um compilador para uma linguagem fictícia no contexto da unidade curricular de Compilação (2024/25). A linguagem MOC adota uma sintaxe semelhante à linguagem C, com restrições e simplificações específicas que visam facilitar a construção de um compilador com ferramentas como **ANTLR4**, não nos vamos alongar na especificação que nos foi dada para a linguagem, mas aqui [link](https://elearning.uab.pt/pluginfile.php/3918150/mod_assign/introattachment/0/MOCC.pdf?forcedownload=1) do enunciado para contextualização:

Por motivos de compatibilidade e possíveis diferenças de comportamento entre versões, recomendamos ANTLR 4.13.2 e Python 3.1**X**, com as quais todos os testes foram realizados.

Sendo o foco desta atividade, a análise léxica e sintática, foram gerados os seguintes ficheiros:

* **main.py** – Ficheiro principal onde está o nosso menu que corre os comandos antlr4 necessários para compilar e verificar o código com as nossas customizações.
* **MOC.g4** – Ficheiro de gramática, inclui regras léxicas(tokens) e sintáticas(parser).
* **MOCErrorListener.py** – Ficheiro de erros customizáveis, traduzindo os erros do parser e lexer para erros mais “user-friendly”.
* **MOCVisitorDEBUG.py** – Ficheiro Visitor customizado com prints a cada momento, que criámos inicialmente para encontrar os problemas e aperfeiçoar a nossa gramática, embora não esteja correntemente a ser usado no nosso main.py.
* **reset\_antlr.sh** – Script auxiliar que criámos para limpar os ficheiros gerados pelo antrl4 e voltar a criar. Usado cada vez que temos de modificar a gramática.
* **README.md** – Instruções de uso e informação técnica do programa.

|  |
| --- |
| ############# Instruções de uso python3 main.py exemplo.txt // Valida a gramática  python3 main.py exemplo.txt -tree // Valida a gramática e gera a árvore sintática  python3 main.py exemplo.txt -gui // Valida a gramática e gera a árvore sintática com interface gráfica |

**Gramática**

**Começámos a definição da gramática pela análise léxica, através da implementação de tokens de gramática antlr, sendo a base que permite construir os elementos básios da nossa linguagem, segue aqui a sua definição:**

* **Palavras-chave**: *int, double, void, main, read, readc, reads, write, writec, writev, writes, if, else, while, for, return,*
* **Operadores**: *+, -, \*, /, %, <, <=, >, >=, ==, !=, &&, ||, !, =,*
* **Pontuação e símbolos**: *,, ;, [, ], {, }, (, ),*
* **Literais**: *números inteiros, números reais e strings,*
* **Identificadores**: *nomes de variáveis, funções, etc.*
* **Comentários**: *comentários de linha (//) e bloco (/\* ... \*/) – ignorados pelo lexer,*
* **Espaços em branco**: ignorados pelo lexer.

A seguir foram implementadas regras sintáxicas, através da implementação de regras de parser que definem se a estrutura dos programas é válida, definidas pelas seguintes regras:

* **programa**: estrutura global que combina os protótipos e o corpo principal do programa.
* **funcao, prototipo, funcaoPrincipal, prototipoPrincipal**: usas para definir e declarar funções, incluindo main.
* **declaracao, variavel:** regras para declaração de variáveis simples e vetores, com ou sem inicialização.
* **instrucao:** regra base para todas as instruções possíveis, incluindo blocos, condições, ciclos, leitura/escrita e retorno.
* **expressao:** criada para suportar operações, nomeadamente: aritméticas, lógicas, relacionais e também operações de casting, chamadas a funções e acessos a vetores.
* **chamadaFuncao:** criada para suportar funções comargumentos: suporte para chamadas como read(), fact(n) e leitura de strings.

Esta informação está toda disponível e comentada corretamente no ficheiro MOC.g4.

Funcionamento do Parser:

O ponto de entrada do parser é a regra programa, que é responsável pela definição da estrutura principal da linguagem definida por esta gramática. Esta regra obriga a que o programa seja composto por dois elementos obrigatórios, um dos elementos é o protótipo e o outro o corpo, terminando com o fim do ficheiro. A ideia de ter esta regra é garantir que os protótipos, sejam obrigatoriamente declaradas antes das funções.

Especificando um pouco o funcionamento dos protótipos, aqui a ideia foi garantir que, esteja garantido pelo menos o protótipo da função main, via prototipoPrincipal, podendo haver antes ou depois declarações dos outros protótipos.

Na regra corpo, podemos ter, nenhuma ou várias unidades, unidades essas que podem ser declarações de variáveis ou de definições de funções com corpo, porém esta regra obriga a que, mesmo que não tenhamos nenhuma unidade declarada, temos de ter a função principal, main.

Na definição dos protótipos e das funções identificamos que parametros estes podem receber, feito com recurso à regra parametros. Com esta regra conseguimos definir os vários tipos de cabeçalhos que podemos ter, ou seja, podemos ter funções que não têm argumentos, ou ter por exemplo a possibilidade de passar vetores como argumentos. A regra parâmetro pode suportar parametros com ou sem identificadores assim como vetores, enquanto a regra tipo, tal como o nome sugere, restringe os tipos suportados pela linguagem, int, double e void.

Na definição das funções é também necessário prever a declaração das variáveis no corpo do programa. Então com recurso à regra declaracao é permitido declarar uma ou mais variáveis dos tipos indicados acima. As variáveis são agrupadas numa listaVariaveis, onde cada elemento é definido pela regra variável, permitindo a flexibilidade para diferentes formas de declaração, podemos ter uma simples variável com identificador, até declarações que são inicializadas, vetores com ou sem valores atribuídos.

Para garantir que estas inicializações e outras operações sejam possíveis, é feita a analise das expressões para permitir a construção de instruções envolvendo operações aritméticas, lógicas, comparações, chamadas de função, acessos a vetores, e conversões de tipo (castings). Dividida por tipos para garantir a presidência e retirar a ambiguidade.

A base destas expressões é construída por elementos como literais numéricos, identificadores e estruturas como chamadas a funções ou acessos a vetores. Para permitir funções com múltiplos argumentos, a regra argumentos aceita uma lista de expressões separadas por vírgulas.

Para garantir as funções de leitura é feita com recurso á regra chamadaFuncao, que define chamadas sem argumentos e com sintaxe fixa.

Seguindo para as instruções que compõe o corpo das funções, começando na regra bloco, que impõe a obrigatoriedade de ter instruções dentro de chavetas. As instruções em si são organizadas por regras específicas, nomeadamente para garantir os ciclos if-else, while, for, etc. Foi feito um esforço para retirar a ambiguidade que é comum existir nos ciclos if-else, separando as instruções if emparelhadas e por emparelhar.

Durante o desenvolvimento da gramática, surgiram alguns desafios específicos que exigiram ajustes manuais, como por exemplo:

* Algumas palavras, como **main**, **read** e **writes**, causaram conflitos na gramática devido a serem identificadores válidos, e embora estivessem definidas como tokens, não eram reconhecidas corretamente pelo parser em algumas regras. Foram utilizadas literalmente nas regras do parser ('main', 'read', etc.), o que resolveu momentaneamente o nosso problema. Mais tarde, após uma análise teórica, compreendemos que a origem do conflito residia na posição incorreta dos tokens no lexer, o que gerava ambiguidade com outras regras, nomeadamente com o token **identificador**. A reorganização das declarações de tokens resolveu o problema de forma definitiva.
* As expressões v(i) (chamada de função) e v[i] (acesso a vetor) tinham sido construídas inicialmente como uma única regra, e apresentavam ambiguidade sintáxica, foi necessário **separar explicitamente** em regras distintas (**chamadaGenerica**, **acessoVetor**), garantindo que o parser conseguisse distinguir corretamente cada forma.
* A regra original de **expressao** tratava apenas operações simples e diretas, mas não reconhecia corretamente estruturas mais complexas como: fact(n), v[i] e castings de (int) x, (double) y. Estas construções fazem parte da base da hierarquia das expressões e, por isso, a limitação residia na definição da regra primary, que representa os elementos terminais de uma expressão. Neste caso tivemos de fazer uma total reformulação da regra primary e as suas ramificações, e forma a suportar expressões compostas e encadeadas, garantindo assim que é respeitado a precedência dos operadores e eliminado ambiguidades.
* Existiu alguns problemas com os protótipos. Sabíamos que tínhamos de garantir a obrigatoriedade da declaração dos protótipos antes das funções e isso foi rapidamente conseguido, mas da forma como o estávamos a fazer obrigava a que o protótipo declarado tivesse obrigatoriamente a função declarada também. E no nosso entendimento, e se a linguagem MOCC é uma derivação da linguagem C, onde, mesmo que o protótipo seja declarado, a função não precisa de existir, com excepção da função main, faria sentido recriar essa situação na nossa gramática. Porém, as várias tentativas levavam sempre a falha sempre que a função não existisse, mas o protótipo sim. Depois de algumas tentativas conseguimos contornar esse problema.

**Listener**

**Validação e testes**

**A nossa estratégia de testes teve como objetivo cobrir os vários aspetos possíveis do compilador, começando por validar cada um dos cenários de sucesso (happy path) da gramática identificados no enunciado (declarações e atribuições) de modo unitário para nos ajudar a validar durante a construção. De seguida avançámos para testes de integração, tendo como dados iniciais os exemplos de programas completos providenciados, focando-nos finalmente nos casos de limite e falha.  
Nos Anexos podem encontrar um resumo mais completo dos testes concluídos.**

**Conclusão e comentários**

**Cenas e coisas   
Falar do visitor**

**Copiar .pdfs e indentação perdida**

Andreia Romão (1702430)

Cátia Santos (1702194)

**Referências**

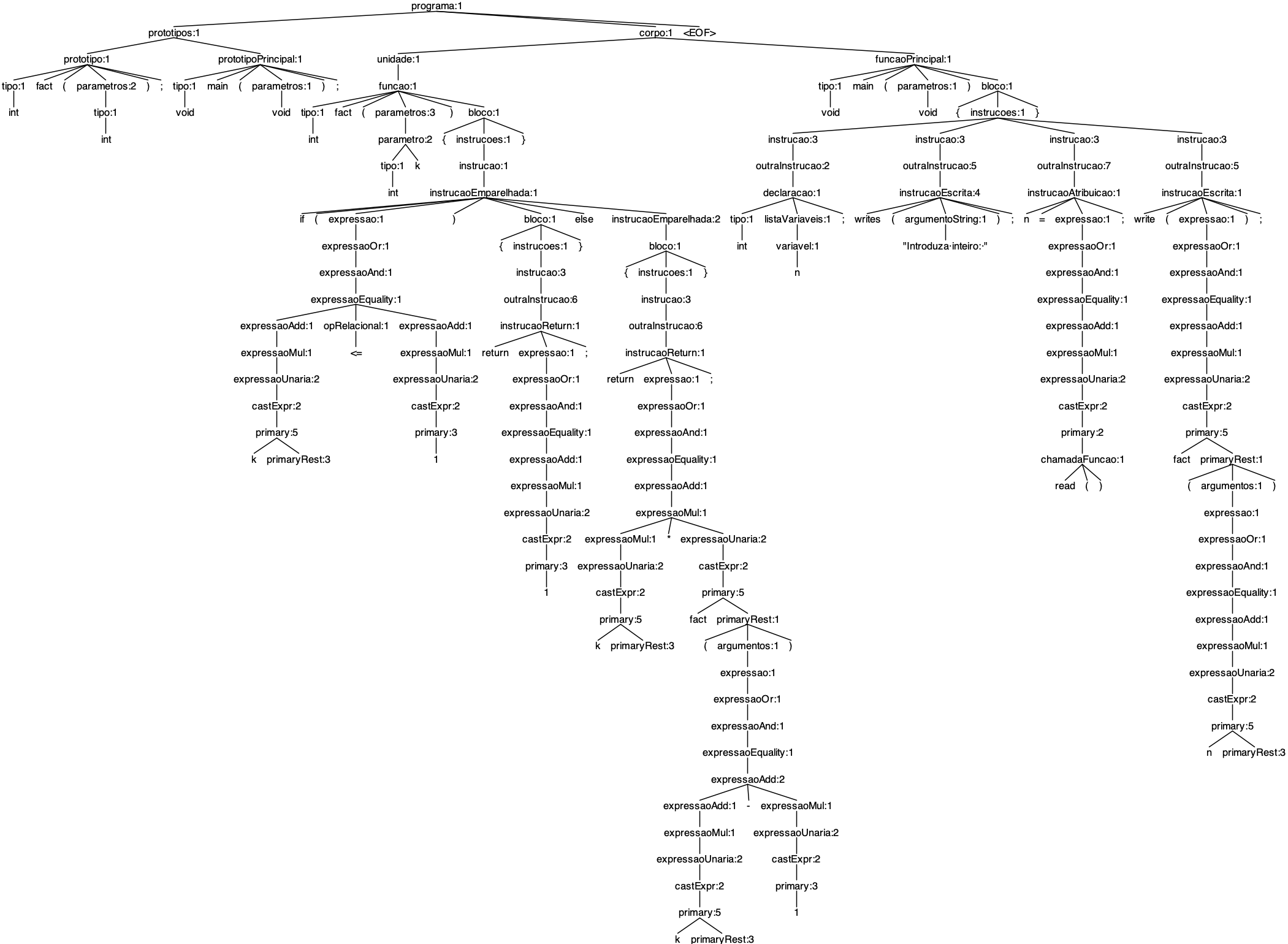
* **Universidade Aberta. (n.d.). Fórum da unidade curricular Compilação [Fórum Moodle]. Acedido através da plataforma Moodle da UAb.**
* **OpenAI. (2025). ChatGPT (versão utilizada para apoio ao desenvolvimento e validação do código).**
* **Adicionar livro do dragão**

**Anexos**

**Execução dos exemplos de sucesso de chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://elearning.uab.pt/pluginfile.php/3959004/mod\_resource/content/1/caso\_sucesso.pdf**

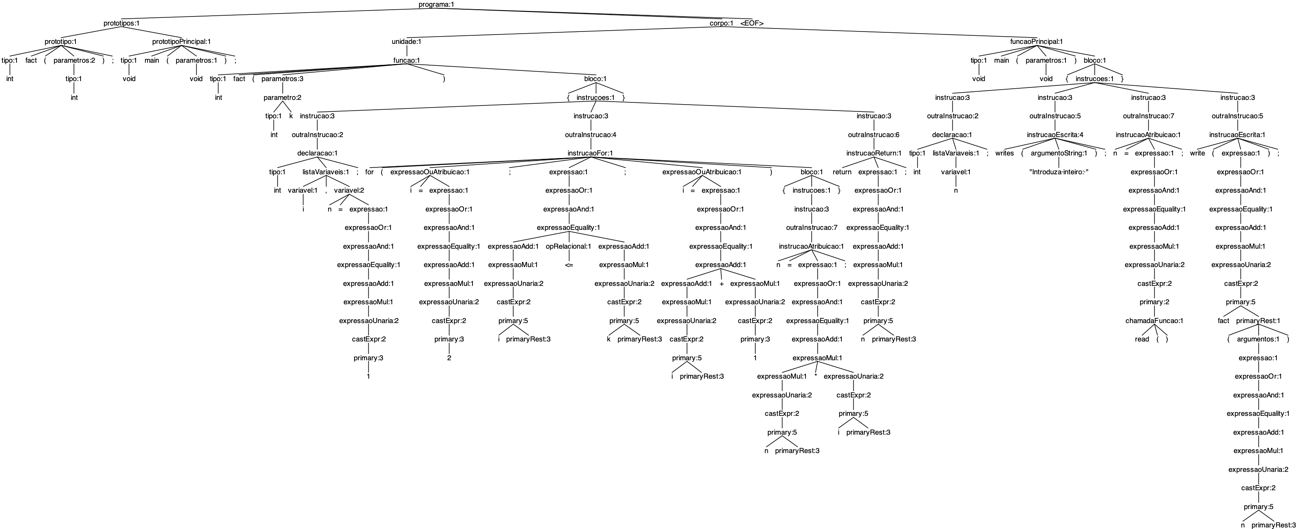
**Exemplo 1 – casos\_sucesso.pdf (corretamente formatado)**

|  |
| --- |
| /\* exemplo 1  fatorial versão recursiva \*/  int fact(int);  void main(void);  int fact(int k) {  if (k <= 1) {  return 1;  } else {  return k \* fact(k - 1);  }  }  void main(void) {  int n;  writes("Introduza inteiro: ");  n = read();  write(fact(n));  } |



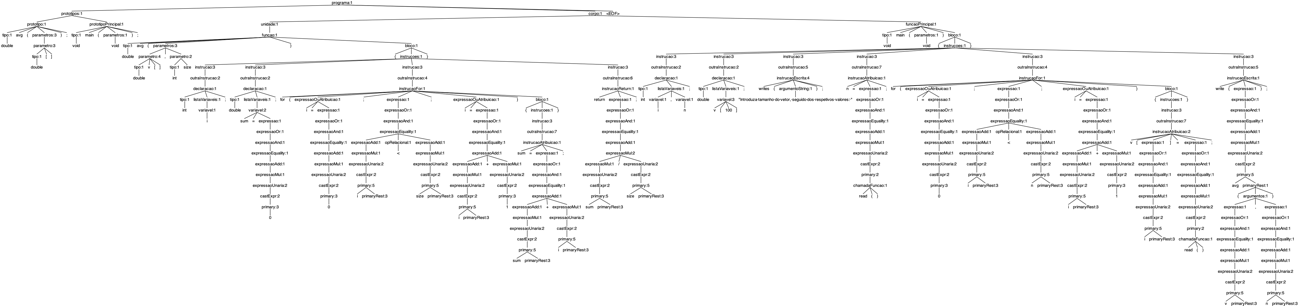
**Exemplo 2 – casos\_sucesso.pdf (corretamente formatado)**

|  |
| --- |
| /\* exemplo 2 \*/  int fact(int);  void main(void);  int fact(int k) {  int i, n = 1;  for (i = 2; i <= k; i = i + 1) {  n = n \* i;  }  return n;  }  void main(void) {  int n;  writes("Introduza inteiro: ");  n = read();  write(fact(n));  } |

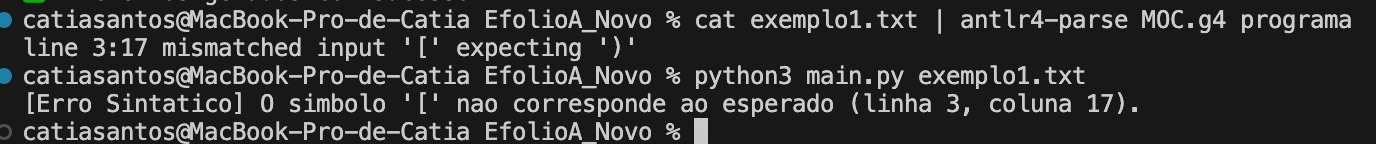


**Exemplo 3 – casos\_sucesso.pdf (corretamente formatado)**

|  |
| --- |
| /\* exemplo 3  média de uma lista de valores positivos \*/  double avg(double[]);  void main(void);  double avg(double v[], int size) {  int i;  double sum = 0;  for (i = 0; i < size; i = i + 1) {  sum = sum + i;  }  return sum / size;  }  void main(void) {  int i, n;  double v[100];  writes("Introduza tamanho do vetor, seguido dos respetivos valores: ");  n = read();  for (i = 0; i < n; i = i + 1) {  v[i] = read();  }  write(avg(v, n));  } |



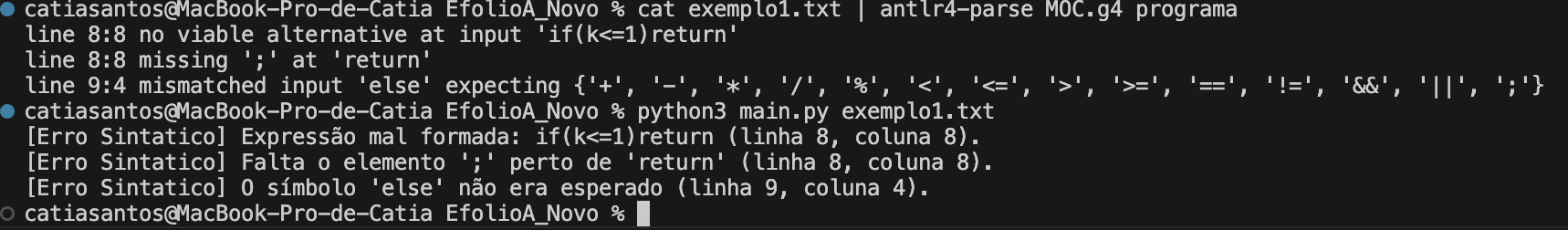
**Já agora a título de mostrar um pouco a funcionalidade do nosso ErrorListener e gramática, de acordo com o problema de protótipos que tínhamos mencionado para este exemplo no Moodle, caso não tivesse sido declaro como parâmetro, este seria o erro do parser vs o erro tratado pelo listener:**

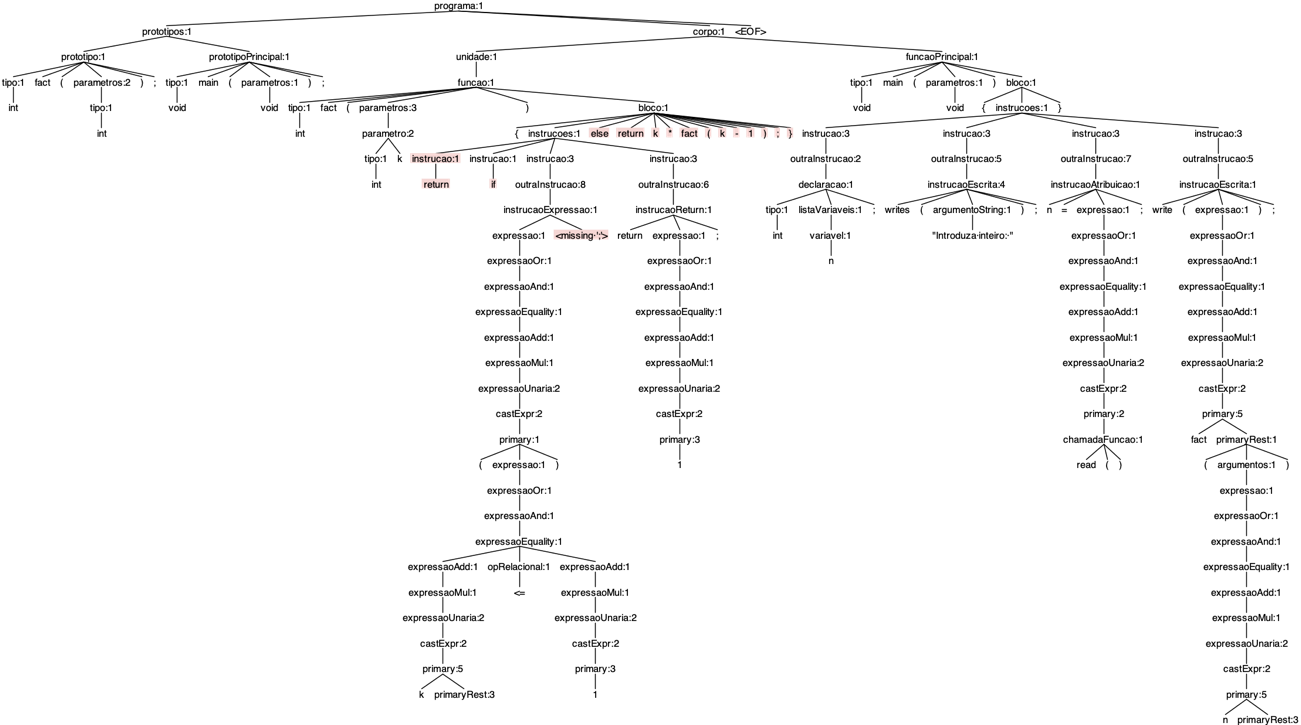


**E agora os testes iniciais do enunciado, testes de falha disponíveis em:** <https://elearning.uab.pt/pluginfile.php/3918150/mod_assign/introattachment/0/MOCC.pdf?forcedownload=1>

**Exemplo 1 – MOOC.pdf (corretamente formatado)**

|  |
| --- |
| /\* exemplo 1  fatorial versão recursiva \*/  int fact(int);  void main(void);  int fact(int k) {  if (k <= 1)  return 1;  else  return k \* fact(k - 1);  }  void main(void) {  int n;  writes("Introduza inteiro: ");  n = read();  write(fact(n));  } |

****



**O Exemplo 2 e Exemplo 3 do MOOC.pdf são os mesmos do casos\_sucesso.pdf , mais acima**

**Seguem também todos os testes unitários executados, divididos em tipo Sucesso (casos de sucesso) e Falha (casos de falha propositada, para ver casos limite), numa tabela com a informação resumida, seguida da execução e árvore para cada exemplo.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **N.ª** | **Descrição** |  | **Tipo** | **Estado** |
| **1.1** | **Teste 1 – Diretivas Pré-processuais**  **Regra: Não pode haver diretivas (ex.: “#include”) no código.** | Sem diretivas | **Sucesso** | **OK** |
| **1.2** | **Contém diretiva proibida “**#include” | **Falha** | **OK** |
| **2.1** | **Teste 2 – Ordem – Protótipos Antes de Declarações**  **Regra: Todos os protótipos devem vir antes de qualquer declaração ou definição.** | Protótipos declarados no início | **Sucesso** | **OK** |
| **2.2** | Variável declarada antes dos protótipos | **Falha** | **OK** |
| **3.1** | **Teste 3 – Declaração de Variáveis: Tipos e Sintaxe**  **Regra: Apenas os tipos "int" e "double" e “void” são permitidos.** | **Uso de tipos permitidos (int, double e void)** | **Sucesso** | **OK** |
| **3.2** | **Uso de tipo não permitido: (float)** | **Falha** | **OK** |
| **4.1** | **Teste 4 – Inicialização de Variáveis**  **Regra: As variáveis podem ser inicializadas com expressões aritméticas ou com valores lidos.** | **Inicialização Aritmética e com leitura** | **Sucesso** | **OK** |
| **4.2** | **Erro de sintaxe: falta parênteses na chamada de read** | **Falha** | **OK** |
| **5.1** | **Teste 5 – Funções de Leitura e Escrita**  **Regra: As funções de leitura devem ser chamadas com parênteses: read(), readc(), reads().**  **Para escrita, usamos: write() para variáveis simples, writev() para vetores, e writes() para strings.** |  | **Sucesso** | **OK** |
| **5.2** | **Fltam parênteses e aspas ausentes em string Usado o write() incorrecto para strings** | **Falha** | **OK** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |
| --- |
| Exemplo de Sucesso 1.1:  int foo(int);  void main(void);  int foo(int a) { return a; }  void main(void) { }  Exemplo de Falha 1.2:  #include <stdio.h> // <-- Isto deve causar erro  int foo(int);  void main(void);  int foo(int a) { return a; }  void main(void) { } |

|  |
| --- |
| Exemplo de Sucesso 2.1:  int foo(int);  void main(void);  int m, n; // <-- Declaração de variáveis após os protótipos  int foo(int a) { return a; }  void main(void) { }  Exemplo de Falha 2.1:  int m, n; // <-- Não é permitido antes dos protótipos  int foo(int);  void main(void);  int foo(int a) { return a; }  void main(void) { } |

|  |
| --- |
| Exemplo de Sucesso 3.1:  int foo(int);  void main(void);  int a, b, v[10];  double x, y, z;  int foo(int a){ return a; }  void main(void){ }  Exemplo de Falha 3.2:  int foo(int);  void main(void);  float a; // <-- 'float' não está definido na gramática  int foo(int a){ return a; }  void main(void){ } |

|  |
| --- |
| Exemplo de Sucesso 4.1:  int foo(int);  void main(void);  int m = 1, n = 2 \* m;  double x = 3.14, y = x / 2;  int a = read();  int foo(int a){ return a; }  void main(void){ }  Exemplo de Falha 4.2:  int foo(int);  void main(void);  int m = 1, n = 2 \* m;  double x = 3.14, y = x / 2;  int a = read; // <-- Falta os parênteses  int foo(int a){ return a; }  void main(void){ } |

|  |
| --- |
| Exemplo de Sucesso 5.1:  int foo(int);  void main(void);  int c = read();  int s[] = reads();  int foo(int a){ return a; }  void main(void){  int v[] = {97,98,99};  write(v[0]); // para número  writec(v[0]); // para carácter  writev(v); // para vetor  writes("abc"); // para string  }  Exemplo de Falha 5.2:  int foo(int);  void main(void);  // Uso incorreto: faltam parênteses e aspas ausentes em string  int c = read; // <-- Falta parênteses  int s[] = reads; // <-- Falta parênteses  int foo(int a){ return a; }  void main(void){  int v[3] = {97,98,99};  write("abc"); // <-- Para escrever string, deve usar writes()  } |