Compte rendu du miniprojet (analyseur syntaxique)

Réalisé par : Ghammam anwar Ichraf ben Fadhel Insaf Khorcheni

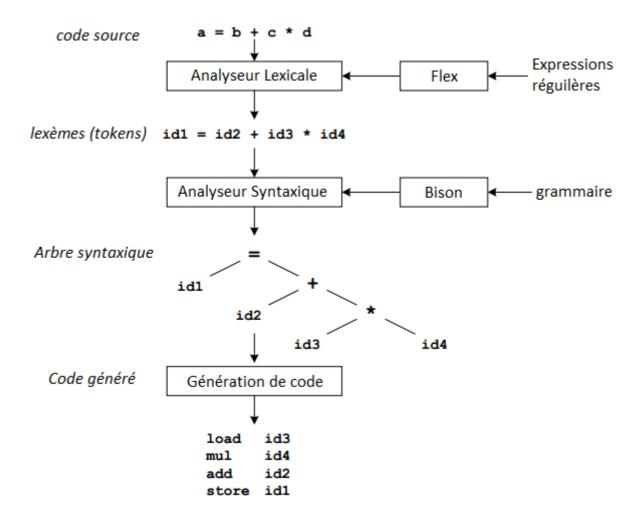
GL4-groupe 2

Introduction

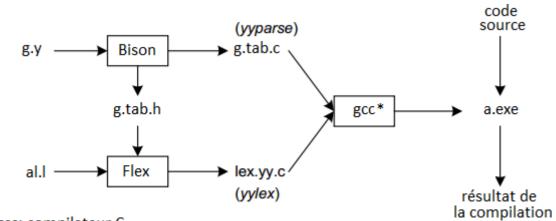
L'intérêt de ce mini-projet est de créer une version simplifié d'un compilateur du langage Pascal.

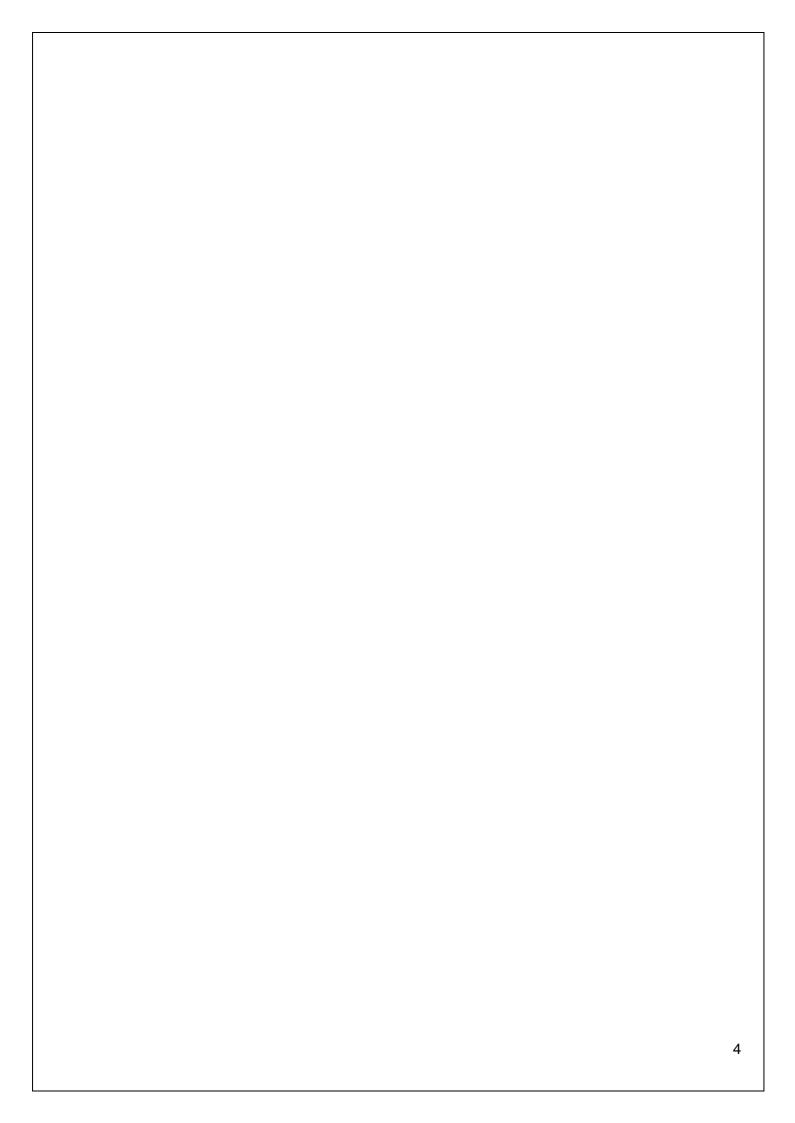
Pour cela, on est amené à créer un analyseur lexicale : pour reconnaître les différents lexèmes (*tokens*) du code source et un analyseur syntaxique : afin de vérifier l'ordre des lexèmes reconnues dans le code.(qui est le but de notre compte rendu)

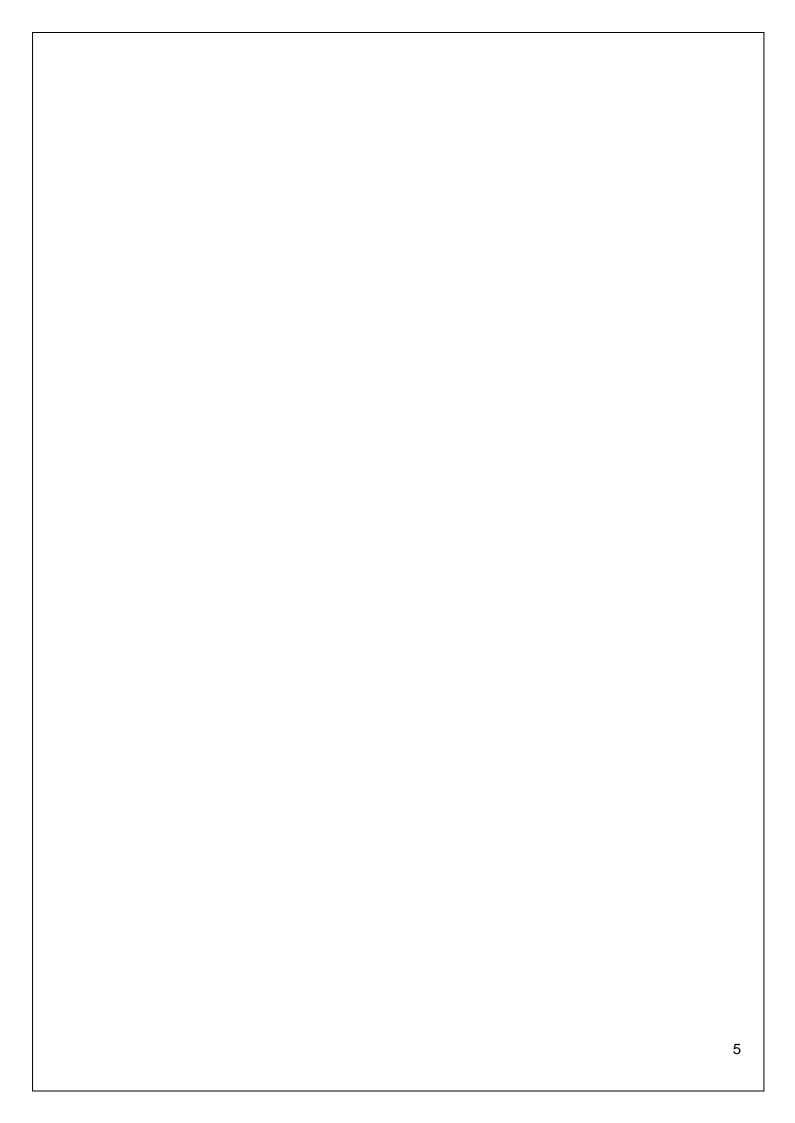
Les outils utilisés sont : bison pour la génération de l'analyseur syntaxique et code::blocks pour compiler les fichiers C générés par flex .



Flex doit obtenir un fichier (.l) pour générer un analyseur lexicale, ce fichier contient les expressions régulières qui définissent les différents lexèmes du langage, bison doit obtenir un fichier (.y) qui contient la description de la grammaire du langage.







Analyse syntaxique

Pour l'analyse syntaxique on doit définir la grammaire qu'utilisent les compilateurs du langage pascal pour vérifier la syntaxe des codes entrés.

Le fichier bison(.y) a le même format que le fichier flex(.l) <u>Définitions</u>: Où on peut inclure les librairies C et déclarer des Variables auxiliaires ou alors des signatures de fonctions à Implémenter dans la dernière partie du fichier.

Définition de la grammaire: Dans cette partie on doit inscrire les Différents règles de production de la grammaire.

```
... définitions ...
%%
... définition de la grammaire ...
... fonctions en C ...
```

Fonctions en C: La dernière partie du fichier bison est consacré à l'implémentation des fonctions principales et optionnels utiles dans la phase de l'analyse sémantique.

```
#include <stdio.h>
int yyerror(char const *msg);
int yylex(void);
extern int yylineno;
   Partie définitions du fichier exemple.y (1/2)
```

```
/* l'axiome de la grammaire */
%token ID
%token NB
%token OPADD
%token OPAFFECT
%token OPMUL
%token OPREL
        /* Mots clés */
%token BEGIN_TOKEN
%token DO
%token ELSE
%token END
%token FUNCTION
%token IF
%token NOT
%token PROCEDURE
%token PROGRAM
%token THEN
%token VAR
%token WHILE
%token INTEGER
%token READ
%token WRITE
%token DEUX_POINTS
%token ARRAY
%token POINT_VIRGULE
%token VIRGULE
%token CHAINE
%token INT
%token PARENTHESE_OUVRANTE
%token PARENTHESE_FERMANTE
%token COMMENTAIRE
```

Partie définitions du fichier pascal.y (2/2)

```
%start programmes
%%
programmes
                                     :PROGRAM
                                      POINT VIRGULE
                                      PROGRAM
                                      ĪD
                                     POINT_VIRGULE
                                      declaration
                                      PROGRAM
                                     POINT_VIRGULE
                                      declaration
                                      instruction_composee
                                                                  { output_syn("fin du programme"); }
                                      PROGRAM
                                      ID
                                      POINT_VIRGULE
                                      instruction_composee
                                                                               {yyerror (" program attendu on line : "); }
{yyerror (" nom du programme invalide on line : "); }
{yyerror (" point virgule attendu on line : "); };
                                      error ID POINT_VIRGULE
                                      PROGRAM error POINT_VIRGULE
                                      PROGRAM ID error
declaration:
         declaration_var
         declarations_sous_programmes
declaration_var:
                                                                                                                                    Activer Windows
         declaration_var
                                                                                                                                    Accédez aux paramètres pour ac
```

Partie définition de la grammaire du fichier pascal.y (1/4)

```
declaration_var:
        declaration_var
        VAR
        liste_identificateurs
        DEUX POINTS
        INTEGER
        POINT_VIRGULE
        /* chaîne vide */
  declaration_var
                { yyerror("mot cle var introuvable !"); }
        liste_identificateurs
        DEUX_POINTS
        INTEGER
        POINT_VIRGULE
        declaration_var
        error { yyerror("identificateur de variable manquante ");}
        DEUX_POINTS
        INTEGER
        POINT_VIRGULE
        declaration_var
    VAR
        liste_identificateurs
        error { yyerror(": manquante !"); }
        INTEGER
        POINT_VIRGULE
```

```
INTEGER
    error { yyerror(" ; manquante !"); }
;
liste_identificateurs:
        ID
        liste_identificateurs
        POINT_VIRGULE
        ID
        liste_identificateurs
        error {yyerror(" ; manquante !");}
        ID
        liste_identificateurs
        POINT_VIRGULE
    error {yyerror(" id introuvable!");}
j
declarations_sous_programmes:
        declarations_sous_programmes
        declarations_sous_programme
        POINT_VIRGULE
        /* chaîne vide */
        declarations_sous_programmes
```

Partie définition de la grammaire du fichier pascal.y (2/4)

```
declarations_sous_programmes:
        declarations_sous_programmes
        declarations_sous_programme
        POINT_VIRGULE
        /* chaÃ⊕ne vide */
        declarations_sous_programmes
        declarations_sous_programme
        error {yyerror(" ; manquante!");}
declarations_sous_programme:
        entete_sous_programme
        declaration
        instruction\_composee
        entete_sous_programme
        instruction_composee
        |error {yyerror(" entete de programme manquante!");}
        declaration
        instruction_composee
        |entete_sous_programme
        declaration
        error {yyerror(" corps de fonction / procedure manquant manquante!");}
;
```

```
entete_sous_programme:
       FUNCTION
       ID
       arguments
       DEUX_POINTS
       INTEGER
       POINT_VIRGULE
       error {yyerror(" mot cle function manquant !");}
       ID
       arguments
       DEUX_POINTS
       INTEGER
       POINT_VIRGULE
       FUNCTION
       error {yyerror(" nom de fonction manquant !");}
       arguments
       DEUX_POINTS
       INTEGER
       POINT_VIRGULE
       FUNCTION
       ID
       arguments
       error {yyerror(" : manquant !");}
       INTEGER
       POINT_VIRGULE
       FUNCTION
```

```
FUNCTION
    ID
    arguments
    DEUX_POINTS
    error {yyerror(" return type de fonction manquant !");}
    POINT_VIRGULE
FUNCTION
    ID
    arguments
   DEUX_POINTS
    INTEGER
    error {yyerror(" ;manquante!");}
    PROCEDURE
    ID
    arguments
    POINT_VIRGULE
    PROCEDURE
    error {yyerror("nom e procedure manquant !");}
    arguments
    POINT_VIRGULE
    PROCEDURE
    ID
    arguments
    error {yyerror(";manquante!");}
```

```
arguments:
       PARENTHESE_OUVRANTE
       liste_parametres
       PARENTHESE_FERMANTE
       error{ yyerror("parenthese ouvrante manquante"); }
       liste parametres
       PARENTHESE_FERMANTE
       PARENTHESE_OUVRANTE
       liste_parametres
       error{ yyerror("parenthese fermante manquante"); }
       /* chaîne vide */
liste_parametres:
       parametre
       liste_parametres
       POINT_VIRGULE
       parametre
       liste_parametres
       error{ yyerror("point virgule manquante"); }
       parametre
error{ yyerror("parametre manquante"); }
       POINT_VIRGULE
       parametre
       liste_parametres
       POINT_VIRGULE
       error{ yyerror("parametre manquante"); }
```

```
parametre:
        ID
       DEUX POINTS
        INTEGER
error{ yyerror("identifiant manquant"); }
        DEUX_POINTS
        INTEGER
ID
        DEUX_POINTS
        error{ yyerror("integer manquant"); }
ID
       DEUX_POINTS
        INTEGER
instruction_composee:
        BEGIN_TOKEN
        instructions_optionnelles
        error{ yyerror("Begin manquante"); }
        instructions_optionnelles
        END
BEGIN_TOKEN
        instructions_optionnelles
        error{ yyerror("End manquante"); }
j
instructions_optionnelles:
        liste_instructions
;
```

```
liste_instructions:
        instruction POINT_VIRGULE
        liste_instructions
        instruction POINT_VIRGULE
        instruction
                                                { yyerror("point virgule d'instruction omis"); }
        error
instruction:
        variable
        OPAFFECT
        expression
        appel_de_procedure
        instruction_composee
        ΙF
        expression
        THEN
        instruction
        ELSE
        instruction
        error { yyerror("IF manquante"); }
        expression
        THEN
        instruction
        ELSE
        instruction
        ΙF
        expression
        error { yyerror("THEN manquante"); }
        instruction
        ELSE
        instruction
```

```
instruction
        ΙF
        expression
        THEN
        instruction
        error { yyerror("ELSE manquante"); }
        instruction
        WHILE
        expression
       DO
        instruction
        /* chaîne vide */
variable:
        ID
appel_de_procedure:
        ID
        ID
        error { yyerror("PARENTHESE OUVRANTE manquante"); }
        liste_expressions
        PARENTHESE_FERMANTE
        ID
        PARENTHESE_OUVRANTE
        liste_expressions
        PARENTHESE_FERMANTE
        PARENTHESE_OUVRANTE
        liste_expressions
        error { yyerror("PARENTHESE_FERMANTE manquante"); }
```

```
liste_expressions:
       expression
liste_expressions
       VIRGULE
       expression
liste_expressions
       error { yyerror("virgule manquante"); }
       expression
П
       liste expressions
       VIRGULE
       error { yyerror("expression manquante"); }
j
expression:
       expression_simple
ı
       expression_simple
       OPREL
       expression_simple
error { yyerror("expression simple manquante"); }
       OPREL
       expression_simple
expression_simple
       error { yyerror("OPREL manquant"); }
       expression_simple
expression_simple
       OPREL
       error { yyerror("expression simple manquante"); }
j
expression_simple:
       terme
       signe
```

```
expression_simple:
        terme
        signe
        error { yyerror("terme manquant"); }
        error { yyerror("signe manquant"); }
        terme
        expression_simple
        OPADD
        terme
        error { yyerror("expression simple manquante"); }
        OPADD
        terme
expression_simple
        error { yyerror("OPADD manquante"); }
        terme
expression_simple
        OPADD
        error { yyerror("terme manquante"); }
;
   terme:
           facteur
           terme
           OPMUL
           facteur
           error { yyerror("terme manquant"); }
           OPMUL
           facteur
           terme
           error { yyerror("OPMUL manquant"); }
           facteur
           terme
           OPMUL
           error { yyerror("facteur manquant"); }
   j
```

```
facteur:
       ID
       ID
       PARENTHESE OUVRANTE
       liste expressions
       PARENTHESE_FERMANTE
ID
       error { yyerror("parenthese ouvrante manquante"); }
       liste expressions
       PARENTHESE_FERMANTE
PARENTHESE_OUVRANTE
       liste expressions
       error { yyerror("parenthese fermante manquante"); }
       PARENTHESE_OUVRANTE
       PARENTHESE_FERMANTE
       NB
       PARENTHESE_OUVRANTE
       expression
       PARENTHESE_FERMANTE
       error { yyerror("parenthese ouvrante manquante"); }
       expression
ı
       PARENTHESE_OUVRANTE
       expression
       error { yyerror("parenthese fermante manquante"); }
       PARENTHESE_OUVRANTE
       error { yyerror("expression manquante"); }
       PARENTHESE FERMANTE
               PARENTHESE_OUVRANTE
               error { yyerror("expression manquante"); }
               PARENTHESE FERMANTE
               NOT
               facteur
               CHAINE
       signe:
               '+'
       %%
```

Partie définition de la grammaire du fichier pascal.y (3/4)

```
int yyerror(char const *msg) {
          fprintf(stderr, "%s %d\n", msg,yylineno);
          return 0;

}
extern FILE *yyin;
int main()
{
        yyparse();
        return 1;
}
```

Partie fonctions en C du fichier pascal.y

<u>NB:</u> la ligne (#include "lex.yy.c") permet d'inclure le fichier (.c) généré par flex pour qu'il soit compilé avec le fichier (.c) généré par bison, plus précisément, le fichier lex.yy.c contient la fonction yylex() vu dans la partie de l'analyse lexicale qui lit le texte en entrée caractère par caractère, reconnait et retourne une suite d'unités lexicales, cette fonction sera appelée par la fonction yyparse() de l'analyseur syntaxique pour obtenir l'unité lexicale suivante. On peut remarquer que la signature de la fonction yylex() se trouve dans la première partie du fichier bisonalors que son implémentation est dans le fichier lex.yy.c

Ace niveau, on doit apporter quelques modifications sur le fichier *flex* pour permettre la coopération entre l'analyseur lexicale et l'analyseur syntaxique: l'option—d de *bison* permet de générer un fichier (.h) contenant les différents symboles terminales codés chacun par un nombre afin de les identifier, ce fichier doit être inclus dans le fichier spécification de l'analyseur lexicale, ensuite, à chaque identification d'une unité lexicale, on doit retourner son code correspondant pour qu'il soit reconnu par *bison*, et si le symbole terminal est écrit à la main (entre guillemets), dans le fichier (.l) on doit retourner le caractère lui-même.

```
#define ID 258
#define NB 259
  efine OPADD 260
  efine OPAFFECT 261
   fine OPMUL 262
   fine OPREL 263
   fine BEGIN 264
   fine DO 265
   fine ELSE 266
   fine END 267
   fine FUNCTION 268
   fine NOT 270
   fine PROCEDURE 271
   fine PROGRAM 272
   fine THEN 273
   fine VAR 274
#define WHILE 275
#define INT 276
```

Les différents codes de symboles terminaux dans le fichier pascal.tab.h

```
/* Définitions */

%{

    /* Code C */
int outputLexical = 1; // flag qui permet d'afficher(1)/masquer(0)
    /* fonction qui affiche un message passé en paramétre */
void output(const char* msg) {
    if(outputLexical == 1) {
        printf("Analyseur lexicale: %s\n", msg);
    }
}

/* chaine de caractéres utilisé avec la fonction output() */
char buffer[50];

#include "pascal.tab.h"
```

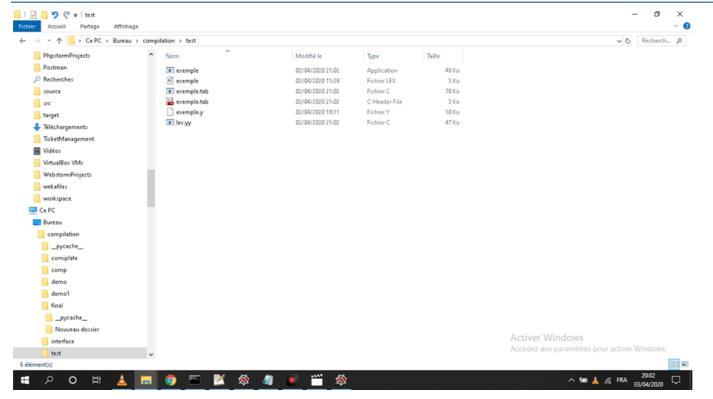
Partie définitions du fichier (./) après modification (Inclusion du fichier pascal.tab.h)

Partie définitions et expressions régulières après modification

```
{nb}
                                              sprintf(buffer, "NB: %s (%d caractere(s))", yytext, yyleng);
                                              output(buffer);
                                              return NB;
{id}
                                               sprintf(buffer, "ID: %s (%d caractere(s))", yytext, yyleng);
                                               output(buffer);
                                               return ID;
==|<>|<|>|<|>|<=|>=
                                             sprintf(buffer, "OPREL: %s (%d caractere(s))", yytext, yyleng);
                                             output(buffer);
                                             return OPREL;
[nN][o0][tT]
                                           sprintf(buffer, "OPREL/NOT: %s (%d caractere(s))", yytext, yyleng);
                                           output(buffer);
                                           return NOT;
```

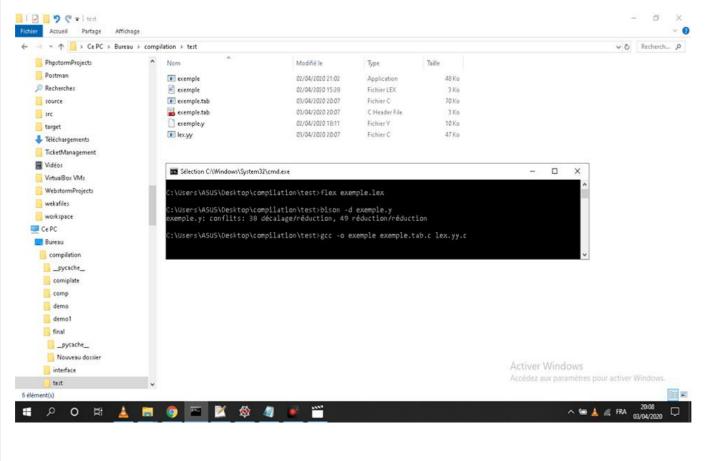
```
[00][rR]
                                              sprintf(buffer, "OPADD: %s (%d caractere(s))", yytext, yyleng);
                                              output(buffer);
                                              return OPADD;
                                             sprintf(buffer, "OPAFFECT: %s (%d caractere(s))", yytext, yyleng);
                                             output(buffer);
                                             return OPAFFECT;
    [dD][iI][vV]
[mM][oO][dD]
    [aA][nN][dD]
                                              sprintf(buffer, "OPMUL: %s (%d caractere(s))", yytext, yyleng);
                                              output(buffer);
                                              return OPMUL;
{blanc}
{commentaire}
                                            sprintf(buffer, "COMMENTAIRE: %s (%d caractere(s))", yytext, yyleng);
                                           output(buffer);
```

les fichiers générés:



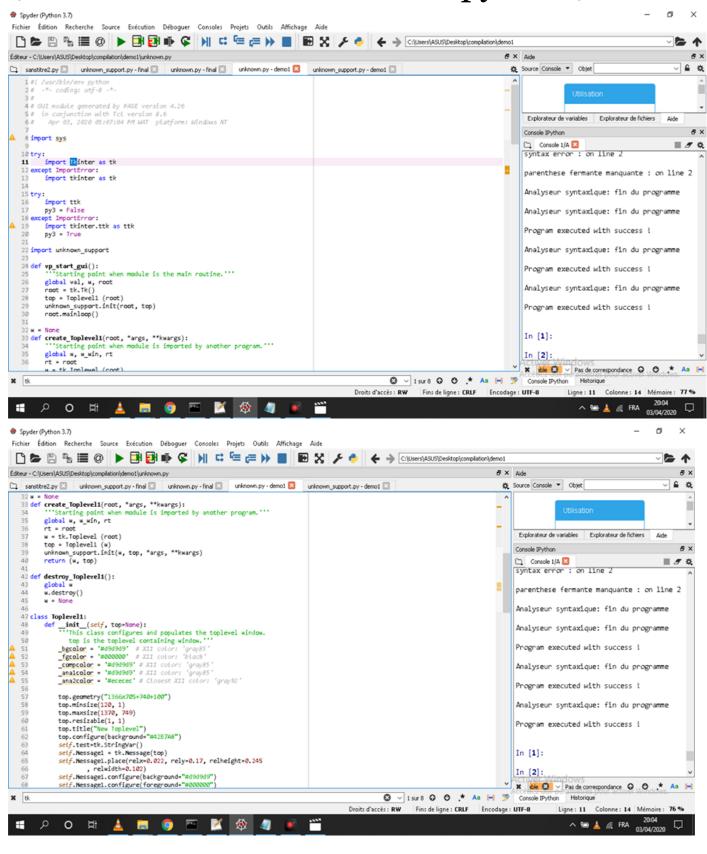
Ceci les etapes pour générer le fichier exemple

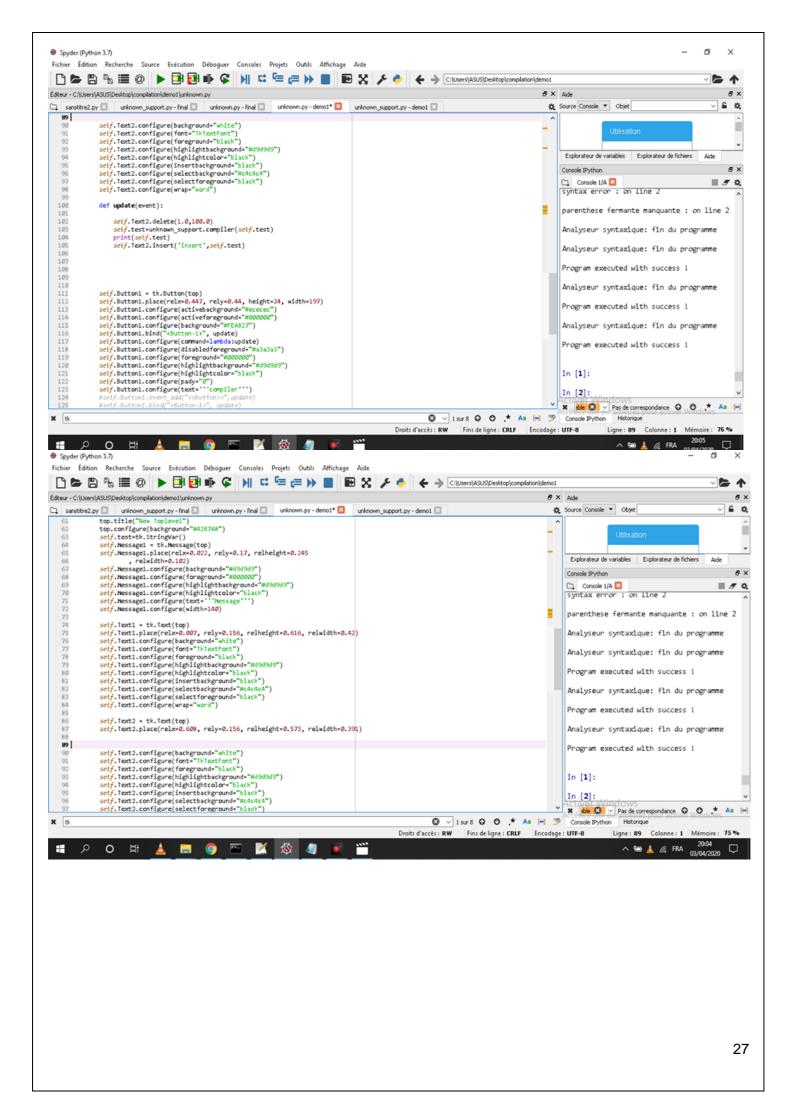
.exe:



fichiers codes de l'interface:

(notre interface a éte codé en python):





codes et leurs résultats de compilations: Un code qui est correcte : programme pascal : program HelloWorld; var i: integed; begin writeln('Hello World'); Le meme code mais cette fois ci erroné : programme pascal : compilation .. program HelloWorld; var i:; begin writeln('Hello World'); end syntax error : on line 3 type de variable manquant 1 : on line 3 Analyseur syntaxique: fin du programme 28

