

Apéndice A

Gramática léxica.

A.1. Átomos.

Para definir las categorías sintácticas de p- (subconjunto de Python) se va a usar BNF extendido donde se abrevia $0 \mid 1 \mid \dots \mid 9$ con $[0-9]$.

Las categorías sintácticas son¹:

- IDENTIFICADOR ::= $([a-zA-Z] \mid -) ([a-zA-Z] \mid [0-9] \mid -)^*$
- BOOLEANO ::= $\text{True} \mid \text{False}$
- ENTERO ::= $[1-9][0-9]^* \mid 0$
- REAL ::= $. [0-9]^+ \mid \text{ENTERO} . [0-9]^+ \mid \text{ENTERO} .$
- CADENA ::= $" <\text{cualquier carácter, excepto } \backslash \text{ o } " >^* "$
- PALABRA RESERVADA ::= $\text{and} \mid \text{or} \mid \text{not} \mid \text{for} \mid \text{while}$
 $\mid \text{if} \mid \text{else} \mid \text{elif} \mid \text{print} \mid \text{in}$
- OPERADOR ::= $+ \mid - \mid * \mid ** \mid / \mid // \mid \% \mid > \mid < \mid > = \mid < = \mid == \mid != \mid - = \mid + =$
- SEPARADOR ::= $(\mid) \mid : \mid ; \mid =$

¹https://docs.python.org/2/reference/lexical_analysis.html#line-structure

En p- la indentación es significativa, esto quiere decir que también debemos tener átomos que distingan los niveles de indentación. El nivel de indentación está dado por el número de espacios y tabuladores al inicio de línea. Tomaremos cada tabulador como un número constante de espacios, 4 (el valor de los tabuladores en Python es variable. No seguiremos esa convención). A diferencia de los átomos de las categorías sintácticas anteriores, éstos no dependen únicamente de empatar patrones, también dependen del contexto en el que fueron reconocidos. A continuación los detalles de cada uno:

- **INDENT**: El nivel de indentación de cada línea determinará el bloque al que pertenece.

Se emitirá uno de estos átomos cada que se reconozca el nivel de indentación y éste sea mayor que el de la línea anterior. Eso quiere decir que se ha empezado un nuevo bloque de código.

- **DEDENT**: Se emitirán estos átomos cada que se reconozca el nivel de indentación actual y sea menor que los anteriores. Por cada nivel anterior y mayor al actual, se emitirá un **DEDENT**. También se emitirá un **DEDENT** por cada nivel que haya entre el del 0 y el de la última línea de código.

- **NEWLINE**: Indica el final de una línea (dependiente del sistema operativo; en Linux es `"\n"`). Las líneas que solo contengan espacios en blanco serán ignoradas. Es decir no generarán átomos **NEWLINE**.

De lo anterior puede deducirse que si el nivel de indentación de una línea es igual al de su predecesora, entonces no será emitido ningún átomo de indentación, ya que pertenecen al mismo bloque.

A.2. Comentarios

A pesar de que los comentarios no representan átomos de la gramática indicaremos cómo deben ser manejados. Los comentarios inician con un `#` y terminan con un

salto de línea. Puede haber comentarios después de código en la misma línea. Esto significaría el fin de la línea. Líneas que sólo empiezan con comentario son ignoradas.