

# **Compiler Team Project #1**

## **-Lexical Analyzer**

02 분반 26 조

20175429 유승훈

20170807 박민기

- overall procedure

일괄된 흐름을 가진 하나의 큰 오토마타로 lexical analyzer 를 구현하였습니다.

초기 구상은 nfa 상의 node 들을 전부 하나의 오토마타에 표기하려 했으나, 해당 방식으로 프로그램을 작성하고 dfa table 을 작성하는 것에 어려움을 느껴 여러 개의 오토마타로 나누어 동작하도록 하였습니다.

처음 input 으로부터 공통된 위치(오토마타: 1, 상태: T0)에서 시작하여, 이후 각각의 오토마타로 process 를 전달시켜 구문 분석을 진행합니다.

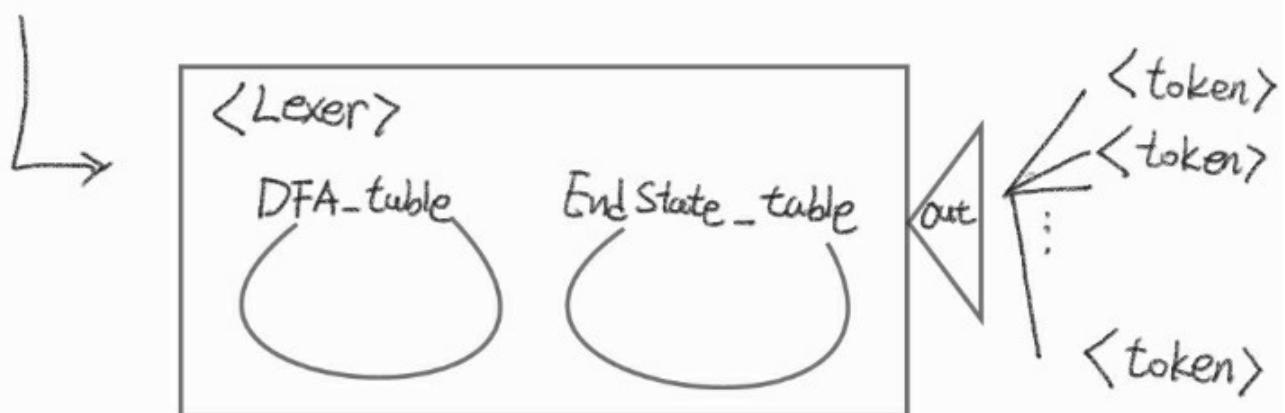
현재 위치에서 입력된 input 을 통해 다음 이동할 수 있는 위치를 dfa\_table 이라는 3 차원 pair<int,int> 배열에 저장해 두었습니다. 다음 위치로 이동이 제한될 경우를 dfa\_table[automata][state] == {0,0} 으로 특정하였고, 이 경우에 EndState\_table 을 확인하여 정상종료와 비정상 종료를 확인하게 됩니다.

EndState\_table 속에 현재 위치가 정상 종료라고 표시되어 있을 경우(EndState\_table[automata][state] != "") 지금까지 받은 문자열을 하나의 token 으로 구분하여 answer 벡터에 저장합니다.

EndState\_table 속에 현재 위치가 비정상종료라고 표시되어 있을 경우(EndState\_table[automata][state] == "") 분석을 종료하고 에러 메세지를 출력하여 어느 문자열에 문제가 있었는지 출력합니다.

만약 분석이 정상적으로 종료된다면, answer 벡터 속 인자들을 입력을 받은 순서대로 토큰의 타입과 종류를 출력하여 줍니다.

[ input\_String ]



- Data structure

```
typedef pair<int, int> ii;
```

오토마타와 상태 변수를 짹으로 저장하기 위해 새롭게 정의한 타입

```
typedef pair<string, string> ss;
```

토큰 저장시, 토큰의 태입과 Lexeme 을 짹으로 저장하기 위해 새롭게 저장한 타입

```
ii DFA_table[15][39][127];
```

현재 오토마타와 상태, 입력 문자를 통해 다음 오토마타와 상태를 저장하는 3 차원 배열

```
string EndState_table[15][39];
```

문자열 입력 종료시, 현재 오토마타와 상태를 통해 토큰의 태입을 특정할 수 있도록 만든 2 차원 배열

- Algorithms

#### \*symbol '-'의 처리방법

'-'가 입력 됐을 시 오토마타 4 번으로 보내어 분석을 진행합니다. 오토마타를 통해 분석 도중

'-'가 홀로 입력 되었을 경우엔 operator 로, 그렇지 않고 위에 '0'를 제외한 digit 등과 합쳐져서 나왔을 경우에는 이전에 입력된 token 과 연결지어 그 의미가 맞는지 분석합니다.

identifier, signed\_integer, RSparen 등의 뒤에 '-'가 올 경우, 이는 operator - 로 동작할 가능성이 있으므로 이경우 해당 입력을 operator '-'와 positive integer 로 구분하여 입력받습니다.

#### \* identifier + special statement + boolean + Vtype

위의 입력들은 알파벳 소문자나 대문자가 처음 입력으로 들어오는 token 이라는 공통점을 가지고 있습니다.

따라서 위의 입력들을 받았을 때 token table 을 정교하게 설계하여 특정한 의미를 갖는 문자열(special statement , boolean , Vtype)과 정확히 일치하며 끝난다면 해당하는 type 으로 인식합니다.

하지만 그렇지 않고 아무런 의미를 갖지 못하는 입력으로 인식된다면 그것을 identifier 로 인식합니다.

- **Example**

input:

0123456789

output:

<Signed\_integer,0>  
<Signed\_integer,123456789>

input:

-1- was-1

output:

<Signed\_integer,-1>  
<Arithmetic\_op,->  
<Identifier,was>  
<Arithmetic\_op,->  
<Signed\_integer,1>

input:

it was if else white while"if" clase 1-+2 = (5)

output:

<Identifier,it>  
<Identifier,was>  
<special\_if,if>  
<special\_else,else>  
<Identifier,white>  
<special\_while,while>  
<String,"if">  
<Identifier,clase>  
<Signed\_integer,1>  
<Arithmetic\_op,->  
<Arithmetic\_op,+>  
<Signed\_integer,2>  
<Assign\_op,=>  
<Identifier,(5>  
<RSparen,)>

error case

input:

"it is a apple

output:

there can't be token like "it is a apple

input:

'ab'

output:

there can't be token like 'ab'

input:

"hello -=-! on

output:

there can't be token like " hello on

(find the first error and exit)

## [Regular Expression]

ZERO = 0

DIGIT = 1 | 2 | 3 | ... | 8 | 9

LETTER = a | b | C ... | z | A | B | C | ... | Z

BLANK = /t/ | /n/ | blank

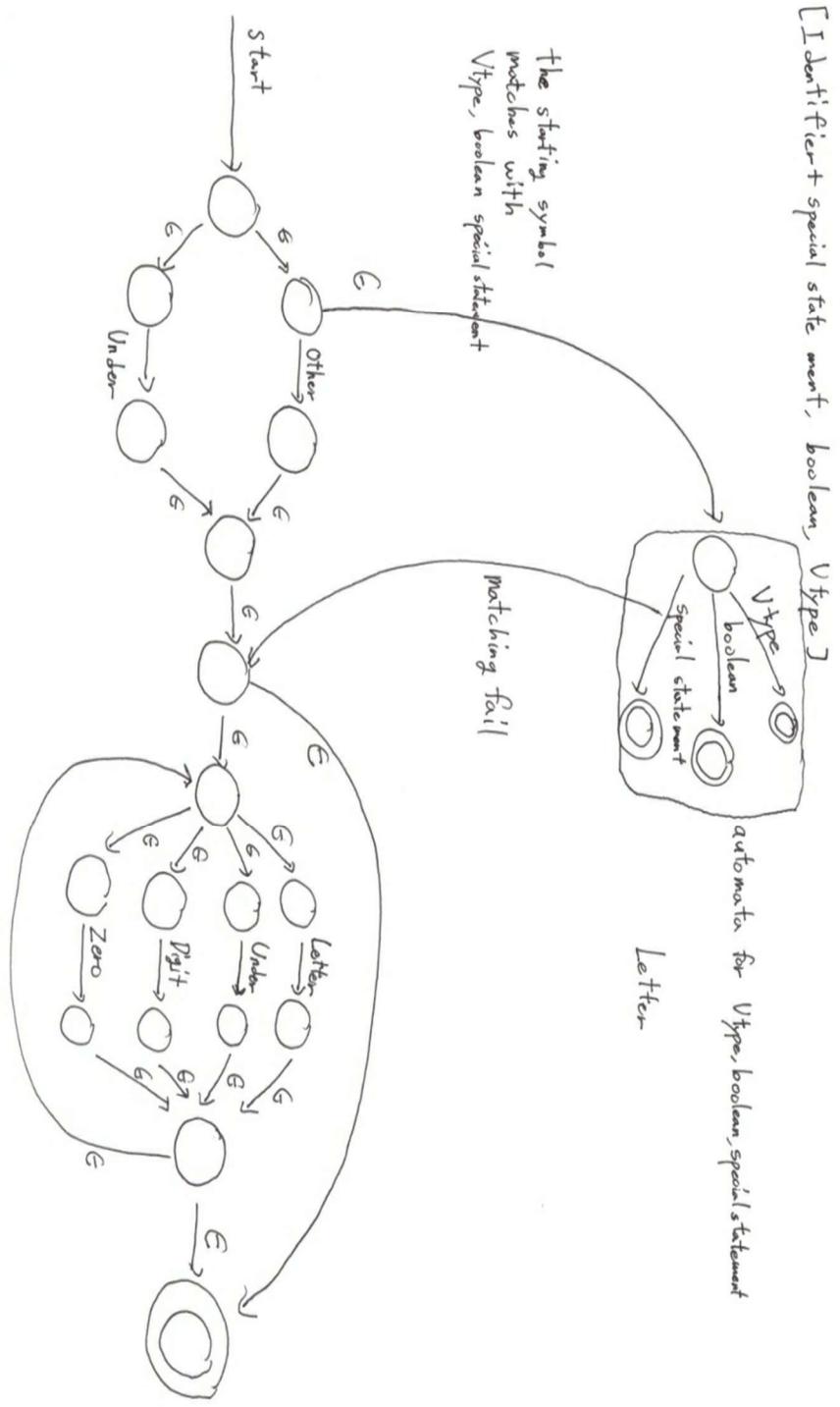
Other = def) Set of characters that are not  
classified in Certain Regular expression

## Single\_Character

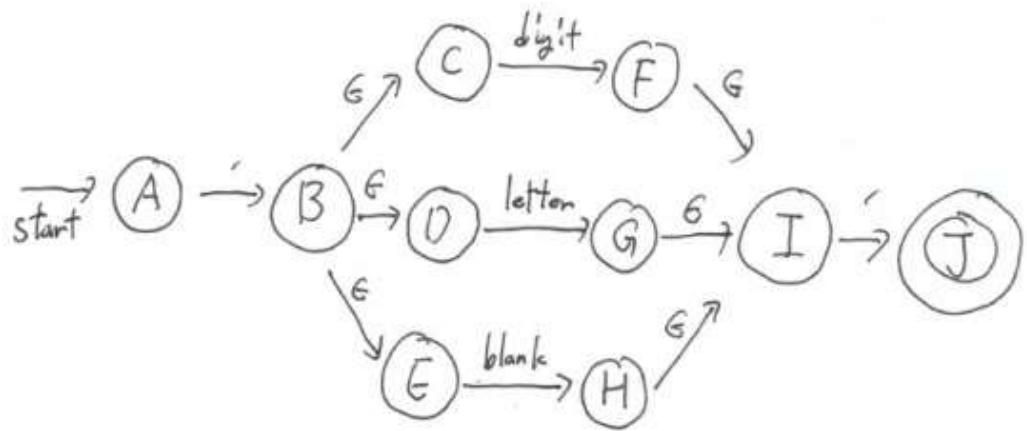
= '( DIGIT | ZERO | BLANK | LETTER )'

## Literal\_String

= "( DIGIT | ZERO | BLANK | LETTER ) ( DIGIT | ZERO | BLANK | LETTER )\*" "



[single character]



$$T_0 = \epsilon\text{-closure}(A) = \{A\}$$

$$T_1 = \epsilon\text{-closure}(\delta(T_0, \cdot)) = \{B, C, D, E\}$$

$$T_2 = \epsilon\text{-closure}(\delta(T_1, \text{digit})) = \epsilon\text{-closure}(F) = \{F, I\}$$

$$T_3 = \epsilon\text{-closure}(\delta(T_1, \text{English letter})) = \epsilon\text{-closure}(G) = \{G, I\}$$

$$T_4 = \epsilon\text{-closure}(\delta(T_1, \text{blank})) = \epsilon\text{-closure}(H) = \{H, I\}$$

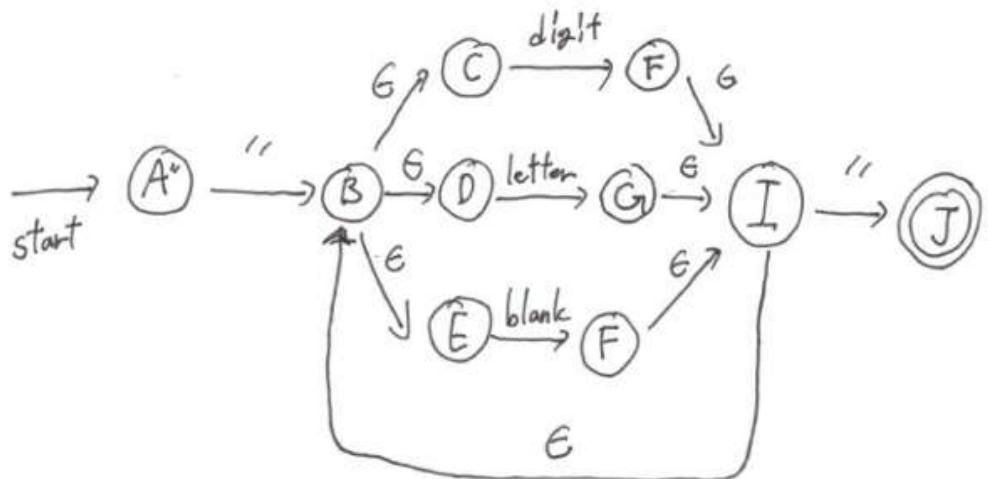
$$T_5 = \epsilon\text{-closure}(\delta(T_2, \cdot)) = \epsilon\text{-closure}(J) = \{J\}$$

$$\epsilon\text{-closure}(\delta([T_1, T_3, T_4, T_5], [\text{digit}, \text{letter}, \text{blank}])) = \emptyset$$

$$\epsilon\text{-closure}(\delta([T_1, T_5], \cdot)) = \emptyset$$

|       | digit       | letter      | blank       | ,           | other       |
|-------|-------------|-------------|-------------|-------------|-------------|
| $T_0$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $T_1$       | $\emptyset$ |
| $T_1$ | $T_2$       | $T_3$       | $T_4$       | $\emptyset$ | $\emptyset$ |
| $T_2$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $T_5$       | $\emptyset$ |
| $T_3$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $T_5$       | $\emptyset$ |
| $T_4$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $T_5$       | $\emptyset$ |
| $T_5$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |

[ Literal string ]



$$T_0 = \text{e-closure}(A) = \{A\}$$

$$T_1 = \text{e-closure}(\delta(T_0, ")) = \{B, C, D, E\}$$

$$T_2 = \text{e-closure}(\delta(T_1, \text{digit})) = \{F, I, B\}$$

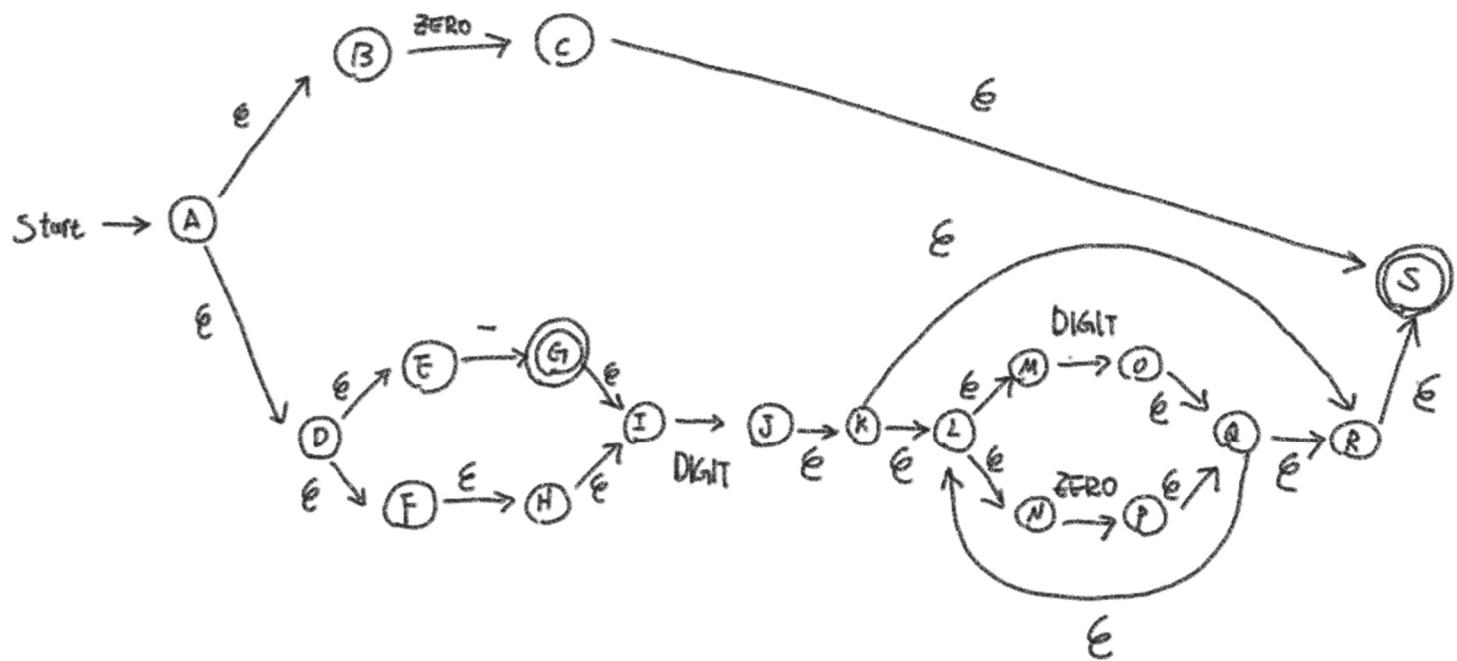
$$T_3 = \text{e-closure}(\delta(T_1, \text{letter})) = \{G, I, B\}$$

$$T_4 = \text{e-closure}(\delta(T_1, \text{blank})) = \{H, I, B\}$$

$$T_5 = \text{e-closure}(\delta(T_3, ")) = \{J\}$$

|       | digit       | letter      | blank       | "           | other       |
|-------|-------------|-------------|-------------|-------------|-------------|
| $T_0$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $T_1$       | $\emptyset$ |
| $T_1$ | $T_2$       | $T_3$       | $T_4$       | $\emptyset$ | $\emptyset$ |
| $T_2$ | $T_2$       | $T_3$       | $T_4$       | $T_5$       | $\emptyset$ |
| $T_3$ | $T_2$       | $T_3$       | $T_4$       | $T_5$       | $\emptyset$ |
| $T_4$ | $T_2$       | $T_3$       | $T_4$       | $T_5$       | $\emptyset$ |
| $T_5$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |

[Signed Integer & operator '-' ]



$\text{INT} = - \mid \text{ZERO} \mid (- \mid \epsilon) \text{ DIGIT } (\text{ZERO} \mid \text{DIGIT})^*$

$$T_0 = \mathcal{E}(A) = \{A, B, D, E, F, H, I\}$$

$$T_1 = \mathcal{E}(\mathcal{G}(T_0, \text{Zero})) = \mathcal{E}(C) = \{C, S\}$$

$$T_2 = \mathcal{E}(\mathcal{G}(T_0, -)) = \mathcal{E}(G) = \{G, I\}$$

$$T_3 = \mathcal{E}(\mathcal{G}(T_0, \text{DIGIT})) = \mathcal{E}(J) = \{J, K, L, M, N, R, S\}$$

$$T_4 = \mathcal{E}(\mathcal{G}(T_3, \text{DIGIT})) = \mathcal{E}(O) = \{O, Q, R, S, L, M, N\}$$

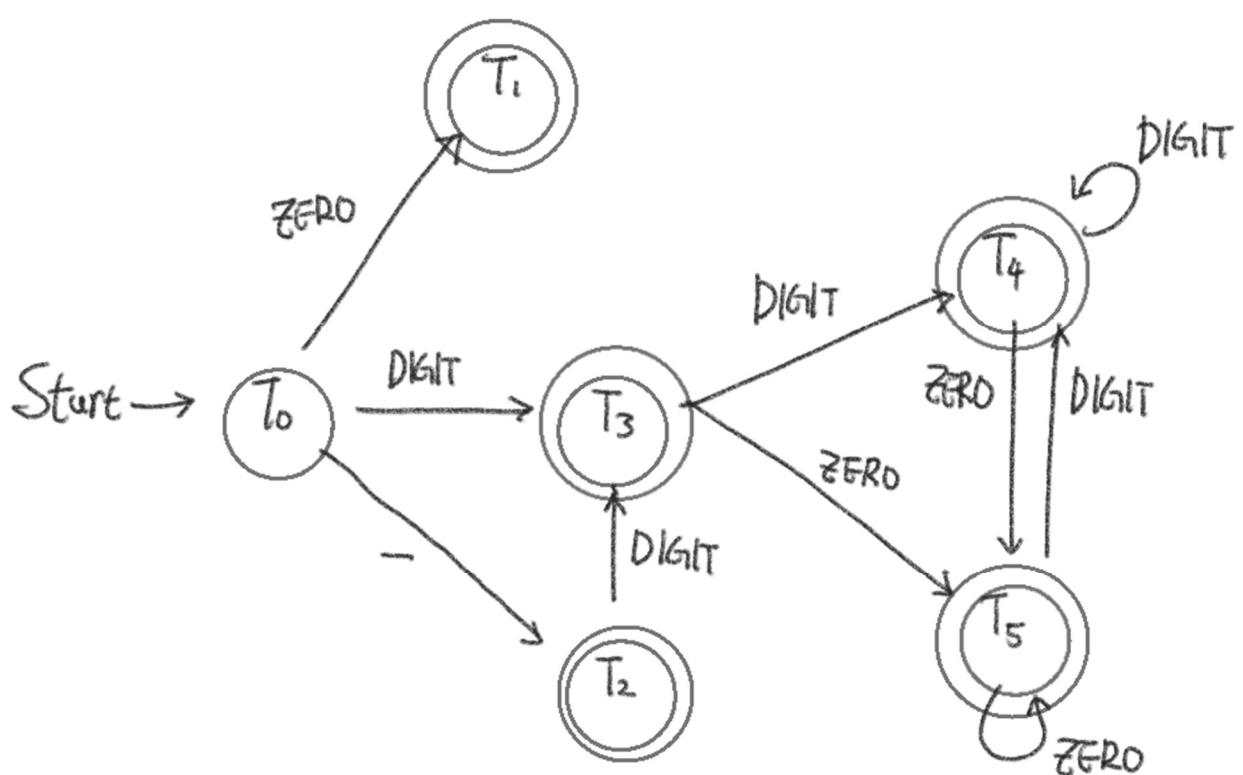
$$T_5 = \mathcal{E}(\mathcal{G}(T_3, \text{ZERO})) = \mathcal{E}(P) = \{P, Q, R, S, L, M, N\}$$

$$\mathcal{E}(\mathcal{G}(T_4, \text{DIGIT})) = T_4 \quad , \quad \mathcal{E}(\mathcal{G}(T_4, \text{ZERO})) = T_5$$

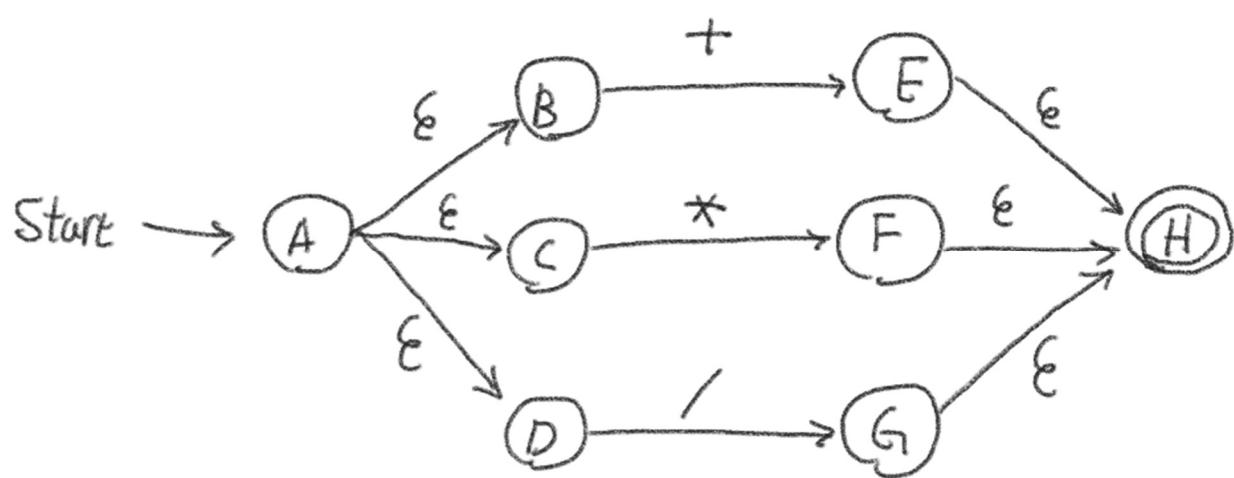
$$\mathcal{E}(\mathcal{G}(T_5, \text{DIGIT})) = T_4 \quad , \quad \mathcal{E}(\mathcal{G}(T_5, \text{ZERO})) = T_5$$

ZERO                    -                    DIGIT                    Other

|       |             |             |             |             |
|-------|-------------|-------------|-------------|-------------|
| $T_0$ | $T_1$       | $T_2$       | $T_3$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |
| $T_2$ | $\emptyset$ | $\emptyset$ | $T_3$       | $\emptyset$ |
| $T_3$ | $T_5$       | $\emptyset$ | $T_4$       | $\emptyset$ |
| $T_4$ | $T_5$       | $\emptyset$ | $T_4$       | $\emptyset$ |
| $T_5$ | $T_5$       | $\emptyset$ | $T_4$       | $\emptyset$ |



# [Arithmetik\_Op]



$$\text{Arithmetik\_Op} = + | * | /$$

$$T_0 = \mathcal{E}(A) = \{A, B, C, D\}$$

$$T_1 = \mathcal{E}(\mathcal{G}(T_0, +)) = \mathcal{E}(E) = \{E, H\}$$

$$T_2 = \mathcal{E}(\mathcal{G}(T_0, *)) = \mathcal{E}(F) = \{F, H\}$$

$$T_3 = \mathcal{E}(\mathcal{G}(T_0, /)) = \mathcal{E}(G) = \{G, H\}$$

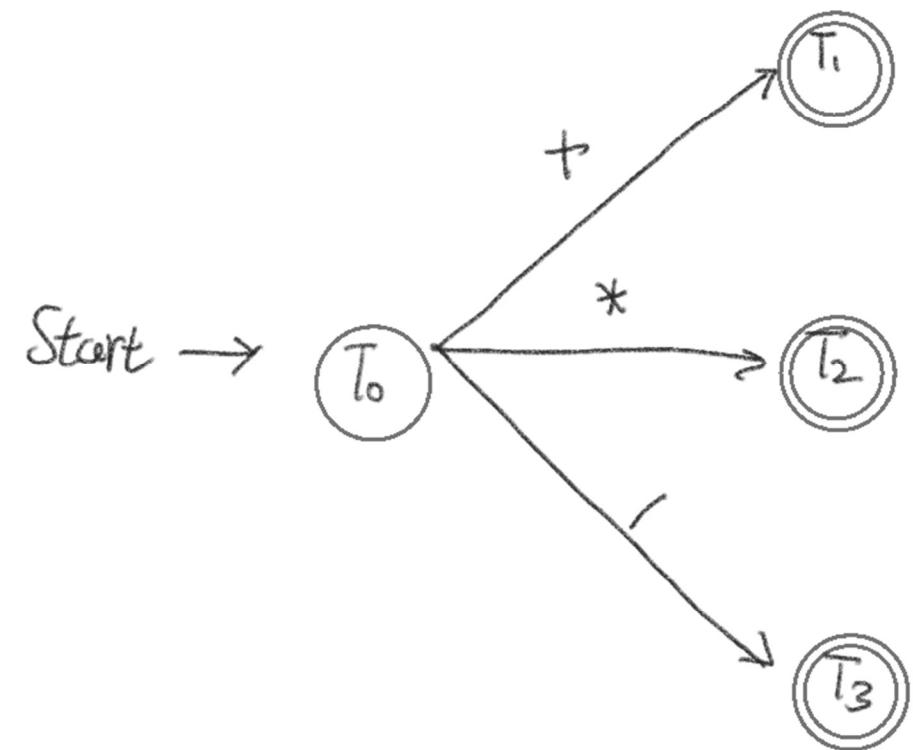
$$\mathcal{E}(\mathcal{G}(T_n, +)) = \emptyset$$

$$\mathcal{E}(\mathcal{G}(T_n, *)) = \emptyset$$

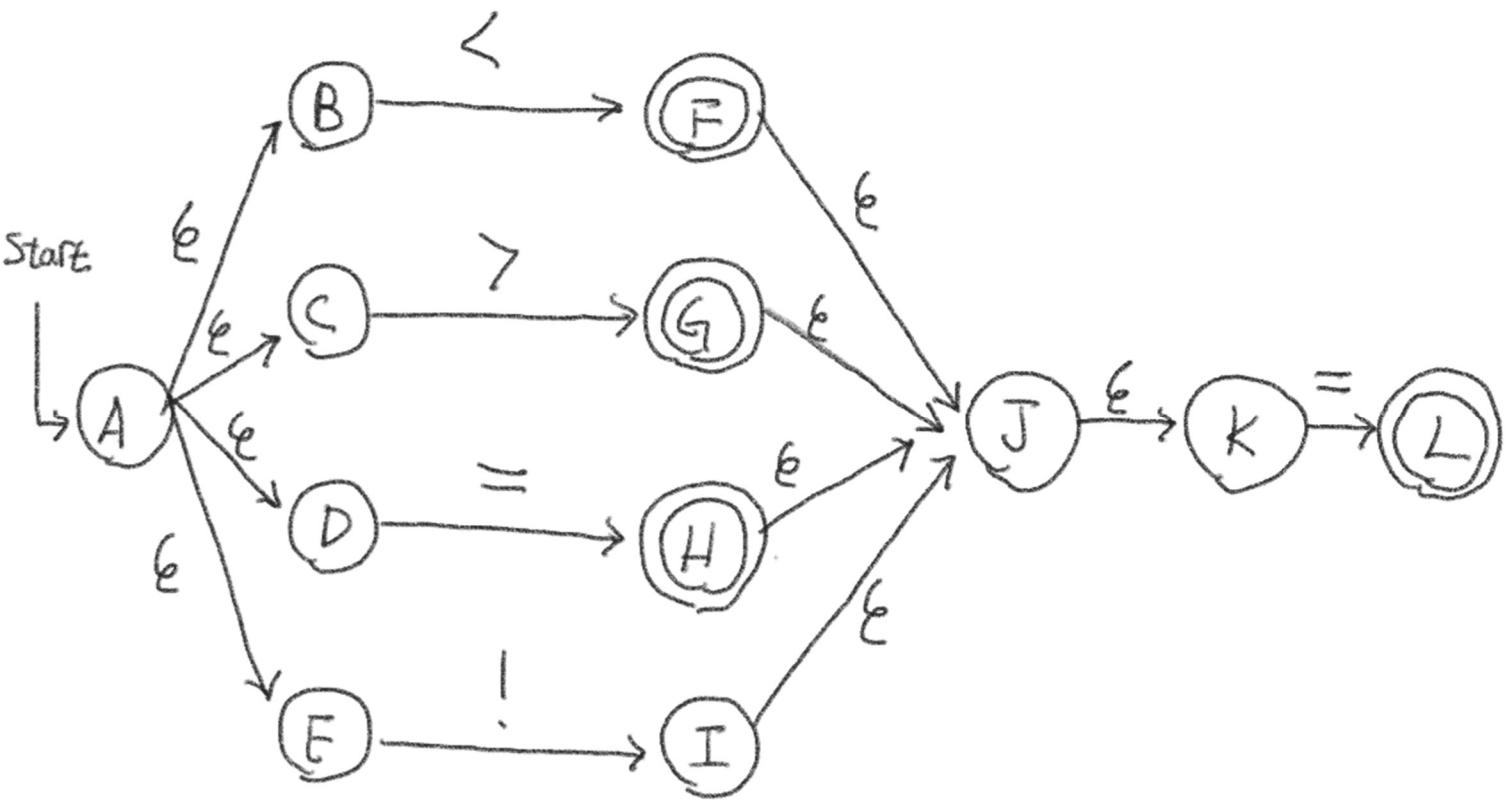
$$\mathcal{E}(\mathcal{G}(T_n, /)) = \emptyset$$

$$*n=1,2,3$$

|       | $t$         | *           | /           | other       |
|-------|-------------|-------------|-------------|-------------|
| $T_0$ | $T_1$       | $T_2$       | $T_3$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |
| $T_2$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |
| $T_3$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |



# [Comparison - Op & Assign - op]



$$OP = != | (< | > | = ) ( = | \epsilon )$$

$$T_0 = \epsilon(A) = \{A, B, C, D, E\}$$

$$T_1 = \epsilon(\mathcal{G}(T_0, <)) = \epsilon(F) = \{F, J, K\}$$

$$T_2 = \epsilon(\mathcal{G}(T_0, >)) = \epsilon(G) = \{G, J, K\}$$

$$T_3 = \epsilon(\mathcal{G}(T_0, =)) = \epsilon(H) = \{H, J, K\}$$

$$T_4 = \epsilon(\mathcal{G}(T_0, !)) = \epsilon(I) = \{I, J, K\}$$

$$T_5 = \mathcal{E}(\overline{\mathcal{E}}(T_n, =)) = \mathcal{E}(L) = \{L\}$$

\*  $n = 1, 2, 3, 4$

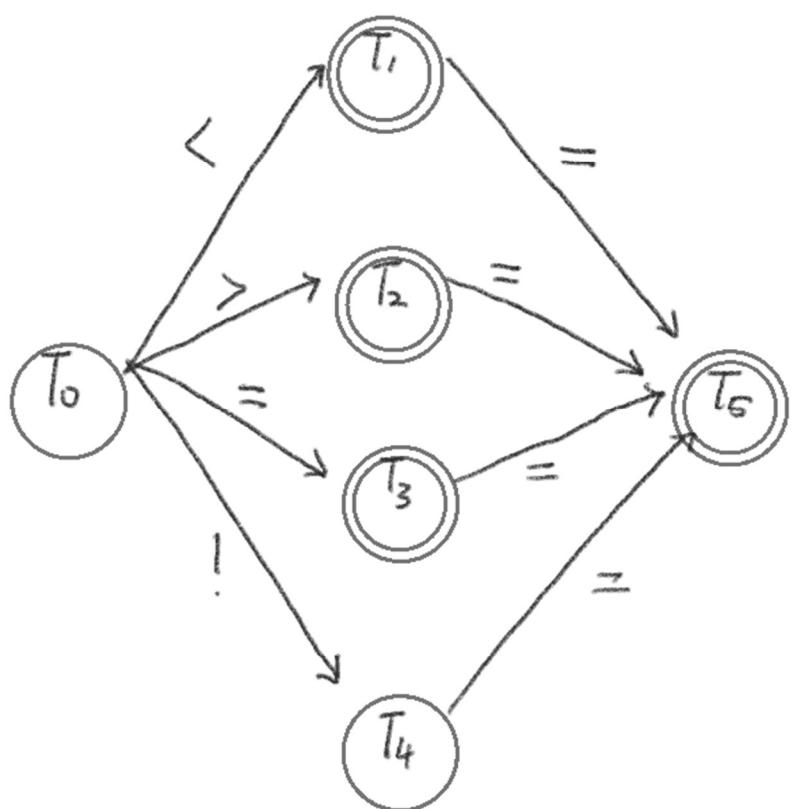
$$\mathcal{E}(\overline{\mathcal{E}}(T_n, <)) = \emptyset$$

$$\mathcal{E}(\overline{\mathcal{E}}(T_n, >)) = \emptyset$$

$$\mathcal{E}(\overline{\mathcal{E}}(T_n, !)) = \emptyset$$

\*  $n = 1, 2, 3, 4, 5$

|       | <           | >           | =           | !           | Other       |
|-------|-------------|-------------|-------------|-------------|-------------|
| $T_0$ | $T_1$       | $T_2$       | $T_3$       | $T_4$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ | $T_5$       | $\emptyset$ | $\emptyset$ |
| $T_2$ | $\emptyset$ | $\emptyset$ | $T_5$       | $\emptyset$ | $\emptyset$ |
| $T_3$ | $\emptyset$ | $\emptyset$ | $T_5$       | $\emptyset$ | $\emptyset$ |
| $T_4$ | $\emptyset$ | $\emptyset$ | $T_5$       | $\emptyset$ | $\emptyset$ |
| $T_5$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |



## [Semi-Colon]

Semi-Colon = ;

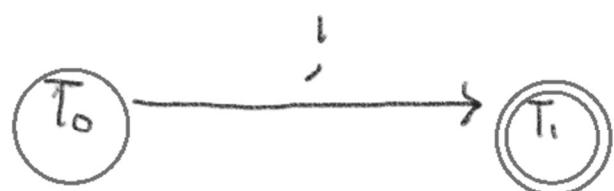


$$T_0 = \epsilon(A) = \{A\}$$

$$T_1 = \epsilon(\bar{\epsilon}(T_0, ;)) = \epsilon(B) = \{B\}$$

$$\epsilon(\bar{\epsilon}(T_1, ;)) = \emptyset$$

|       | ;           | Other       |
|-------|-------------|-------------|
| $T_0$ | $T_1$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ |



[Comma]

Comma = ,



$$T_0 = \mathcal{E}(A) = \{A\}$$

$$T_1 = \mathcal{E}(\mathcal{E}(T_0, ,)) = \{B\} = \{B\}$$

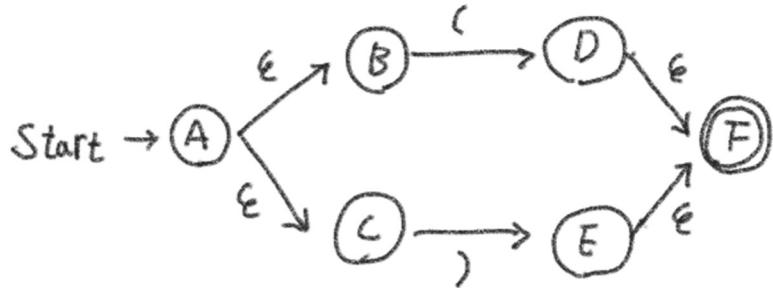
$$\mathcal{E}(\mathcal{E}(T_1, ,)) = \emptyset$$

|       | ,           | Other       |
|-------|-------------|-------------|
| $T_0$ | $T_1$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ |



## [ Parentheses ]

Paren = ( | )



$$T_0 = \epsilon(A) = \{A, B, C\}$$

$$T_1 = \epsilon(\tilde{\epsilon}(T_0, ())) = \epsilon(D) = \{D, F\}$$

$$T_2 = \epsilon(\tilde{\epsilon}(T_1, ,)) = \epsilon(E) = \{E, F\}$$

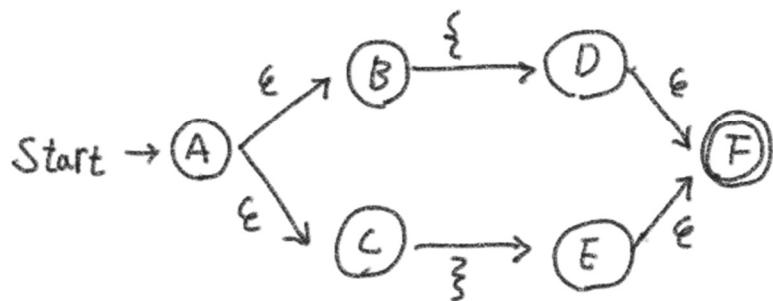
$$\epsilon(\tilde{\epsilon}(T_1, ())) = \epsilon(\tilde{\epsilon}(T_1, ,)) = \emptyset$$

$$\epsilon(\tilde{\epsilon}(T_2, ())) = \epsilon(\tilde{\epsilon}(T_2, ,)) = \emptyset$$

|       | (           | )           | other       |
|-------|-------------|-------------|-------------|
| $T_0$ | $T_1$       | $T_2$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |
| $T_2$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |

[ brace ]

$$\text{brace} = \{ \mid \}$$



$$T_0 = \mathcal{E}(A) = \{A, B, C\}$$

$$T_1 = \mathcal{E}(\mathcal{E}(T_0, \{ \})) = \mathcal{E}(D) = \{D, F\}$$

$$T_2 = \mathcal{E}(\mathcal{E}(T_1, \{ \})) = \mathcal{E}(E) = \{E, F\}$$

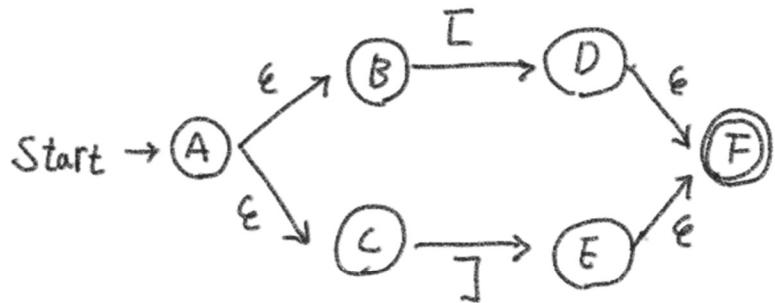
$$\mathcal{E}(\mathcal{E}(T_1, \{ \})) = \mathcal{E}(\mathcal{E}(T_1, \})) = \emptyset$$

$$\mathcal{E}(\mathcal{E}(T_2, \{ \})) = \mathcal{E}(\mathcal{E}(T_2, \})) = \emptyset$$

|       | $\{$        | $\}$        | other       |
|-------|-------------|-------------|-------------|
| $T_0$ | $T_1$       | $T_2$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |
| $T_2$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |

[bracket]

bracket = [ | ]



$$T_0 = \mathcal{E}(A) = \{A, B, C\}$$

$$T_1 = \mathcal{E}(\mathcal{G}(T_0, ())) = \mathcal{E}(D) = \{D, F\}$$

$$T_2 = \mathcal{E}(\mathcal{G}(T_1, ())) = \mathcal{E}(E) = \{E, F\}$$

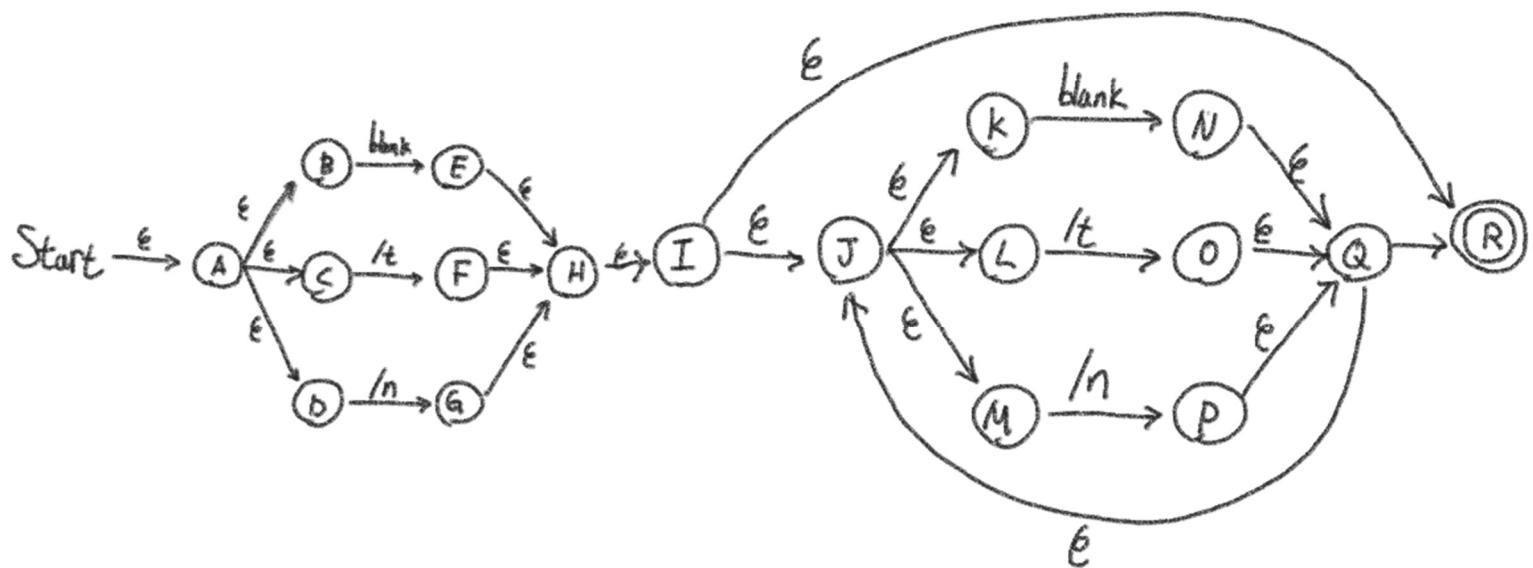
$$\mathcal{E}(\mathcal{G}(T_1, ())) = \mathcal{E}(\mathcal{G}(T_1, ())) = \emptyset$$

$$\mathcal{E}(\mathcal{G}(T_2, ())) = \mathcal{E}(\mathcal{G}(T_2, ())) = \emptyset$$

|       | [           | ]           | other       |
|-------|-------------|-------------|-------------|
| $T_0$ | $T_1$       | $T_2$       | $\emptyset$ |
| $T_1$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |
| $T_2$ | $\emptyset$ | $\emptyset$ | $\emptyset$ |

[ White\_Space ]

$$\text{white\_space} = (\text{blank} \mid \text{t} \mid \text{n}) (\text{blank} \mid \text{t} \mid \text{n})^*$$



$$T_0 = \mathcal{E}(A) = \{A, B, C, D\}$$

$$T_1 = \mathcal{E}(\mathcal{G}(T_0, \text{blank})) = \mathcal{E}(E) = \{E, H, I, J, K, L, M, R\}$$

$$T_2 = \mathcal{E}(\mathcal{G}(T_0, \text{t })) = \mathcal{E}(F) = \{F, H, I, J, K, L, M, R\}$$

$$T_3 = \mathcal{E}(\mathcal{G}(T_0, \text{n })) = \mathcal{E}(G) = \{G, H, I, J, K, L, M, R\}$$

$$T_4 = \mathcal{E}(\mathcal{G}(T_1, \text{blank})) = \mathcal{E}(N) = \{N, Q, J, K, L, M, R\}$$

$$T_5 = \mathcal{E}(\mathcal{G}(T_1, \text{t })) = \mathcal{E}(O) = \{O, Q, J, K, L, M, R\}$$

$$T_6 = \mathcal{E}(\mathcal{G}(T_1, \text{n })) = \mathcal{E}(P) = \{P, Q, J, K, L, M, R\}$$

$$\epsilon(\bar{\epsilon}(T_2, \text{blank})) = \epsilon(w) = T_4, \quad \epsilon(\bar{\epsilon}(T_3, \text{blank})) = \epsilon(w) = T_4$$

$$\epsilon(\bar{\epsilon}(T_2, /t)) = \epsilon(w) = T_5, \quad \epsilon(\bar{\epsilon}(T_3, /t)) = \epsilon(w) = T_5$$

$$\epsilon(\bar{\epsilon}(T_2, /n)) = \epsilon(w) = T_6, \quad \epsilon(\bar{\epsilon}(T_3, /n)) = \epsilon(w) = T_6$$

\* n = 4, 5, 6

$$\epsilon(\bar{\epsilon}(T_n, \text{blank})) = \epsilon(N) = T_4$$

$$\epsilon(\bar{\epsilon}(T_n, /t)) = \epsilon(N) = T_5$$

$$\epsilon(\bar{\epsilon}(T_n, /n)) = \epsilon(N) = T_6$$

|                | blank          | /t             | /n             | Other |
|----------------|----------------|----------------|----------------|-------|
| T <sub>0</sub> | T <sub>1</sub> | T <sub>2</sub> | T <sub>3</sub> | ∅     |
| T <sub>1</sub> | T <sub>4</sub> | T <sub>5</sub> | T <sub>6</sub> | ∅     |
| T <sub>2</sub> | T <sub>4</sub> | T <sub>5</sub> | T <sub>6</sub> | ∅     |
| T <sub>3</sub> | T <sub>4</sub> | T <sub>5</sub> | T <sub>6</sub> | ∅     |
| T <sub>4</sub> | T <sub>4</sub> | T <sub>5</sub> | T <sub>6</sub> | ∅     |
| T <sub>5</sub> | T <sub>4</sub> | T <sub>5</sub> | T <sub>6</sub> | ∅     |
| T <sub>6</sub> | T <sub>4</sub> | T <sub>5</sub> | T <sub>6</sub> | ∅     |