I.3 Tata Bahasa

- Jika diberikan spesifikasi suatu bahasa, ada dua masalah utama yang berkaitan dengan masalah tersebut :
  1. Bagaimana membangkitkan satu atau lebih kalimat didalam bahasa tersebut ?
     - Masalah pembangkitan (*Generation*)
  2. Bagaimana menentukan apakah suatu kalimat termasuk di dalam bahasa tersebut atau tidak?
     - Masalah Pengenalan (*Recognition*)
- Aturan yang digunakan untuk membangkitkan atau mengenali kalimat di dalam suatu bahasa dinamakan tata bahasa(*grammar*).
- **Contoh :** Tata bahasa dibawah ini adalah sebagian dari tata bahasa untuk membangkitkan atau mengenali kalimat di dalam Bahasa Inggris :
  1. Sebuah **sentence** dibentuk dari **noun phrase** diikuti dengan **verb phrase.**
  2. Sebuah **sentence** dibentuk dari **noun phrase** diikuti dengan **verb phrase** dan **noun phrase.**
  3. dll
- Kita dapat menuliskan semua kaidah tata bahasa di atas secara ringkas sebagai berikut :

  <sentence> → <noun phrase> <verb phrase>

  <sentence> → <noun phrase> <verb phrase> <noun phrase>

  <noun phrase> → <adjective><noun phrase>

  <noun phrase> → <adjective><singular noun >

  <verb phrase > → <singular verb> <adverb>

  <verb phrase > → <singular verb>

  <adjective> → *the*

  <adjective> → *a*

  <adjective> → *little*

<singular noun> → *boy*

<singular noun> → *dog*

<singular verb> → *runs*

<singular verb> → *bites*

<adverb> → *quickly*

- Tanda panah (→) pada kaidah di atas menyatakan bahwa item di sebelah kiri panah dapat diganti dengan item di sebelah kanan panah.
- Proses pembangkitan kalimat *the little boy runs quickly* dapat diturunkan sebagai berikut :

  <sentence> → <noun phrase> <verb phrase>

  → <adjective> <noun phrase> <verb phrase>

  → *the* <noun phrase> <verb phrase>

  → *the* <adjective> <singular noun > <verb phrase>

  → *the little* <singular noun > <verb phrase>

  → *the little boy* <verb phrase>

  → *the little boy* <singular verb> <adverb>

  → *the little boy runs* <adverb>

  → *the little boy runs quickly*
- Proses pembangkitan kaliamat diatas juga dapat digambarkan dalam suatu diagram pohon- yang disebut *pohon parsing*.
- Unsur-unsur tata bahasa :

  Secara formal, suatu tata bahasa terdiri dari empat unsur :

  1. Himpunan berhingga **terminal T**
  2. Himpunan berhingga **non –terminal N**
  3. Himpunan berhingga **aturan produksi P**
  4. **Simbol awal S Є N**
- Penjelasan :

  1. Simbol terminal didalam *T* adalah symbol yang dapat digunakan untuk membuat kalimat/string di dalam bahasa itu. Simbol terminal biasanya dilambangkan dengan huruf kecil, seperti a,b,c,… atau angka 0,1,..

Contoh : *T* = {a,b,c}

2. Simbol non terminal di dalam *N* adalah lambang antara yang digunakan untuk mendeskripsikan struktur kalimat. Simbol non-terminal biasanya dilambangkan dengan hruf besar, seperti A,B,C,..
   Contoh : *N* = {S,A,B}
3. Produksi adalah kaidah/aturan tata bahasa yang mengatur cara pembentukan kalimat di dalam bahasa.
   Setiap aturan dituliskan dalam bentuk :

$$\alpha \rightarrow \beta$$

   yang dalam hal ini, **α dan β** adalah string yang dibentuk dari himpunan *T U N*, dan **α** harus berisi paling sedikit satu simbol non-terminal.
4. Simbol awal adalah simbol non-terminal khusus yang mengawali pembangkitan kalimat di dalam bahasa tersebut.

- Keempat unsur tata bahasa tersebut sering dilambangkan dengan :

$$G = (T,N,P,S)$$

  Bahasa yang dihasilkan oleh *G* ditulis sebagai L(*G*), yaitu himpunan *string* yang dapat diturunkan dan simbol awal *S* dengan menerapkan aturan produksi yang terdapat di dalam *P.*

- Contoh :

  Tinjau tata bahasa $G = (T,N,P,S)$, dengan $T=\{a,b\}$, $N=\{S,A\}$, $P=\{S \rightarrow aA, S \rightarrow b, S \rightarrow aa\}$

  Apakah *L(G)* dari Tata bahasa *G* ?

  Dengan mencoba semua aturan produksi dari simbol awal *S* sebagai berikut :

  (i) $S \rightarrow aA$

  $\rightarrow aaa$

  (i) $S \rightarrow b$

  maka hanya terdapat dua *string* yang dapat dibangkitkan, yaitu *b* dan *aaa*, tidak ada yang lain. Jadi *L(G) = {b, aaa}*.