

Estruturas Sintáticas

João Paulo Lazzarini Cyrino

27/10/2020

Aqui estão algumas anotações de aspectos principais do que há no livro de Syntactic Structures e que não foram abordados no conteúdo de linguagens formais.

Argumentos contra a gramática markov

Chomsky propõe o modelo de estrutura sintagmática frente a duas deficiências de se assumir que a gramática é uma máquina de estados finita (modelo markov). A primeira se relaciona com o fato de modelos estatísticos, em tese, não poderem gerar algo como “incolores ideias verdes dormem furiosamente”, uma vez que essa frase – apesar de estar de acordo com a gramática da língua – é formada por bigramas que jamais seriam encontrados em algum corpus: incolores ideias, ideias verdes, verdes dormem, dormem furiosamente.

A segunda deficiência tem relação com tipos de língua que uma gramática markov não é capaz de gerar. Considerando línguas com apenas os símbolos a e b , as línguas abaixo seriam impossíveis se a gramática fosse um modelo markov (máquina de estados finitos):

- $ab, aabb, aaabbb$ - todas as sentenças que contenham n ocorrências de a seguidas por n ocorrências de b e apenas estas sentenças.
- $aa, bb, abba, baab, aaaa, bbbb, aabbaa, abbbba$, e todas as sentenças que consistam de uma sequência X seguida por uma imagem em espelho de X e apenas estas sentenças.
- $aa, bb, abab, aaaa, bbbb, aabaab, abbabb$, e todas as sentenças que consistam de uma sequência X seguida pela mesma sequência X e apenas estas sentenças.

Uma gramática markov não conseguiria gerar essas línguas pois as regras que definem suas sentenças envolvem uma memória maior do que a de um único símbolo, ou a de um número fixo de símbolos. Na primeira, as sentenças dependem de uma regra envolvendo um número n variável de símbolos. Na segunda e na terceira, a regra envolve uma memória de uma sequência X qualquer, de tamanho também variável. Gramáticas markov sempre seguem um modelo de bigramas, tendo apenas memória do símbolo anterior. Mesmo modelos de trigramas, quadrigramas possuem a memória fixa e não variável de 3 e 4 símbolos respectivamente.

Nas línguas naturais encontramos instâncias em que regras como as acima seriam necessárias. Por exemplo frases como:

- Se S_1 , então S_2 .
- Ou S_1 , ou S_2 .
- O homem que disse que S_1 está chegando hoje.

Considere que S_1 e S_2 são sentenças. Nesse caso, para que ocorra *então* depois de *se* S_1 seria necessário que a gramática pudesse acessar o *se* que se encontra antes de S_1 , de tamanho variável. Claramente isso não é possível em um modelo markov. O mesmo vale para a segunda e, de certa forma, para a terceira sentença.

Modelo de Estrutura Sintagmática

A alternativa para o modelo markov encontrada por Chomsky é a gramática de estrutura sintagmática. Essa gramática consiste, basicamente, de símbolos terminais, não-terminais e regras de substituição (reescrita ou

produção). Para uma compreensão da formalização, ver o material sobre linguagens formais.

Sentenças são sequências de símbolos e uma língua é um conjunto de sentenças geradas por uma gramática.

A geração de sentenças se dá da seguinte forma:

1. Toma-se um símbolo raiz (S) e substitua-o por símbolos de acordo com uma das regras de produção.
2. Toma-se os símbolos não-terminais e substitua-os de acordo com uma das regras de produção.
3. A geração da sentença termina até que haja apenas símbolos terminais

Línguas naturais são, normalmente, modeladas pelo que se chama de Gramáticas Livres de Contexto ou *Context-free Grammars (CFG)*. Nessas gramáticas, as regras de substituição assumem que o lado esquerdo contenha apenas um símbolo não-terminal e o lado direito contenha qualquer sequência de símbolos (terminais ou não-terminais).

A regra de produção abaixo ilustra isso. Um símbolo não-terminal raiz (S) pode ser substituído por uma sequência de dois símbolos não-terminais: SN e SV (referem-se a sintagma nominal e sintagma verbal, respectivamente)

$$S \rightarrow SN \cdot SV$$

O Modelo Transformacional

O modelo transformacional é o modelo de Gramática Gerativa lançado por Chomsky em 1957 em Estruturas Sintáticas. Após discorrer sobre a importância de um modelo de estrutura sintagmática, tal como a *gramática livre de contexto* mencionada acima, Chomsky lança mão das **transformações** para aumentar o poder descritivo das *gramáticas livres de contexto*.

Uma *transformação* é um tipo de regra que se aplica a uma sentença gerada por uma gramática gerativa (livre de contexto, no caso) e a transforma em outra sentença. Exemplos são transformar sentenças declarativas em interrogativas, ativas em passivas, etc.

Elas são definidas como no exemplo abaixo:

Regra de voz Passiva

- Análise estrutural: $SN - Aux - V - SN$
- Mudança estrutural: $X_1 - X_2 - X_3 - X_4 \rightarrow X_4 - X_2 + be + en - X_3 - by + X_1$

Transformações serão importantes ao longo da história da Teoria Gerativa. Embora as motivações para essas transformações sejam fracas dentro do modelo nesse momento inicial, elas terão melhores explicações nos modelos seguintes de Teoria Gerativa.