



人工知能特論II 第2回

二宮 崇 1

今日の講義の予定

- CFG (Context Free Grammar, 文脈自由文法)
- 言語学での発展



C F G



CFG (文脈自由文法): 導入

- 正規言語と正規文法 (有限オートマトンが受理する表現とその文法 ; c.f. regular expression)
 - (“a” + “b”)* は “a” と “b” の任意の組み合わせの文字列を受理する。E.g., aabababbab...
- 正規文法の限界
 - $a^n b^n$ だけを受理する文法がつかれない
 - 入れ子表現「(太郎は(花子が使っていた) ペンを借りた)」など
- チョムスキーが言語を扱うために次の文法階層を規定
 - 文脈依存文法 (Context Sensitive Grammar)
 - 文脈自由文法 (Context Free Grammar)

CFG: 生成

CFG
S → SUBJ VP
S → VP
SUBJ → N が
VP → OBJ V
VP → V
OBJ → N を
N → S N
N → N と N
V → 送った
V → 読んだ
N → 香織
N → 恵
N → 電子メール

S ⇒ SUBJ VP
⇒ N が VP
⇒ 香織 が VP
⇒ 香織 が OBJ V
⇒ 香織 が N を V
⇒ 香織 が S N を V
⇒ 香織 が SUBJ VP N を V
⇒ 香織 が N が VP N を V
⇒ 香織 が 恵 が VP N を V
⇒ 香織 が 恵 が V N を V
⇒ 香織 が 恵 が 送った N を V
⇒ 香織 が 恵 が 送った 電子メール を V
⇒ 香織 が 恵 が 送った 電子メール を 読んだ

S^{*} ⇒ 香織 が 恵 が 送った 電子メール を 読んだ₅

CFG: 構文木

CFG

S → SUBJ VP

S → VP

SUBJ → N が

VP → OBJ V

VP → V

OBJ → N を

N → S N

N → N と N

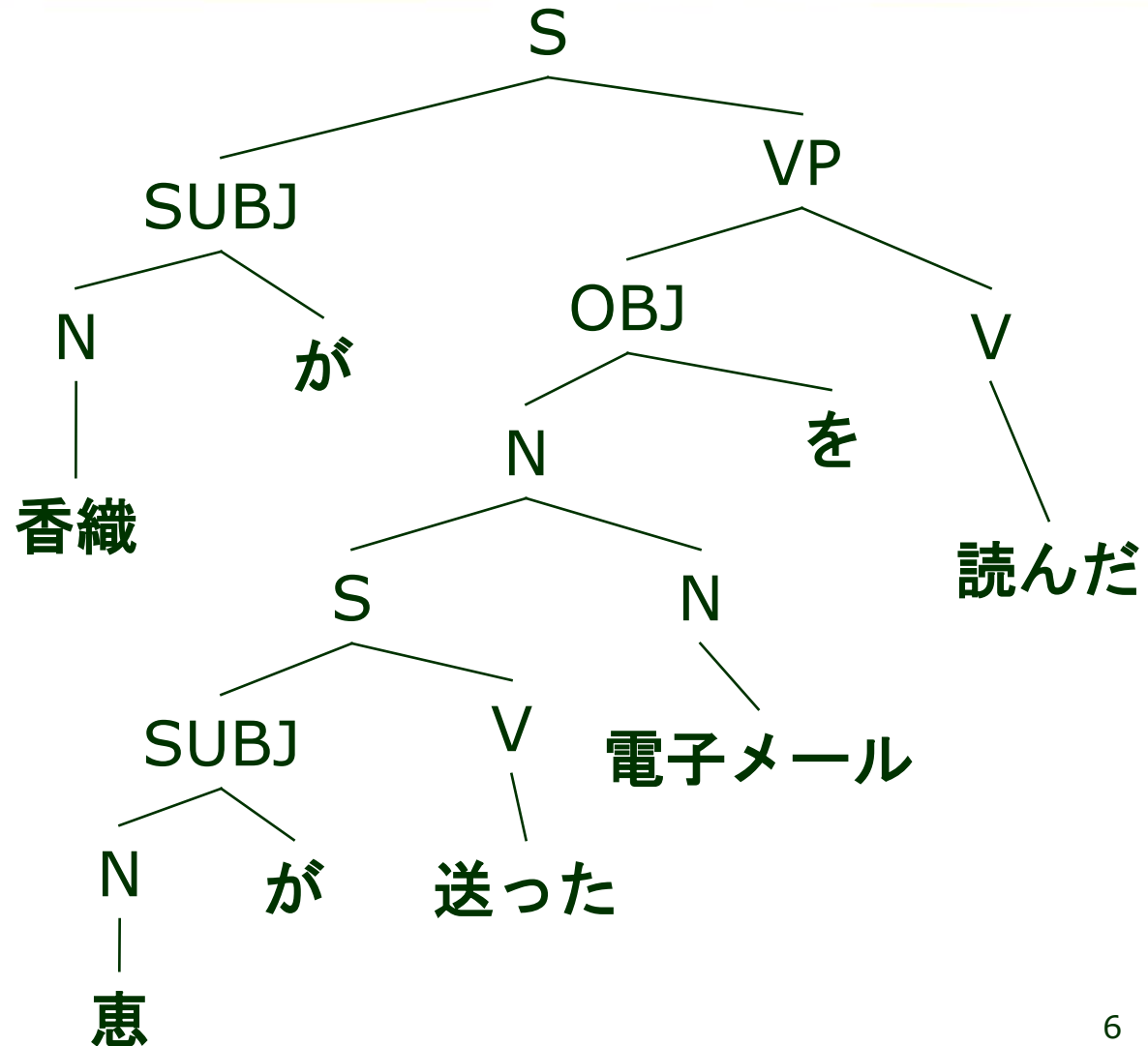
V → 送った

V → 読んだ

N → 香織

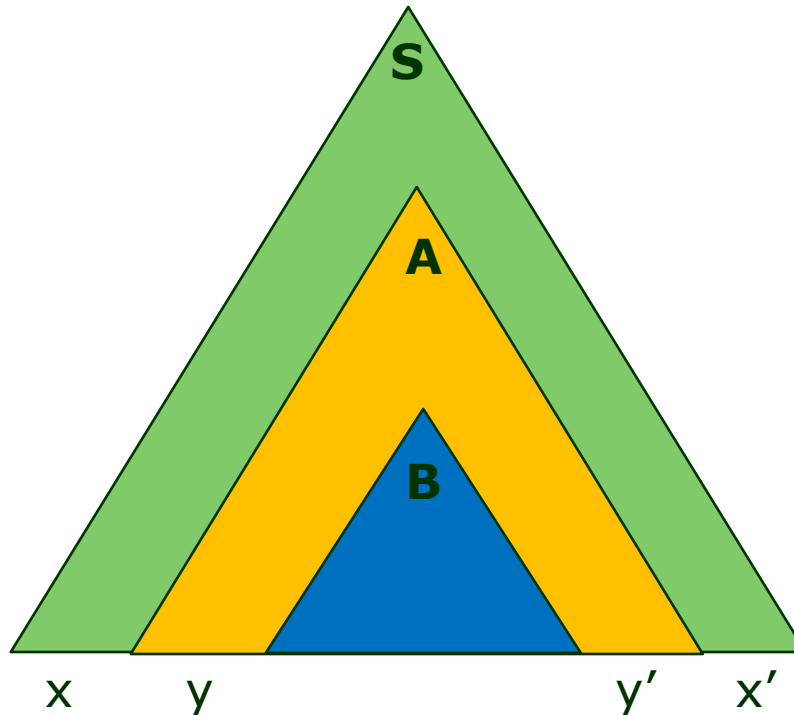
N → 恵

N → 電子メール



CFG: 再帰的

- $NP \rightarrow S NP$
 - 太郎が好きな花子が好きな次郎が好きな恵が好きな香織が好きな.....



形式文法

- $G (= \langle V_N, V_T, P, \sigma \rangle)$
 - V_N : 非終端記号の集合
 - V_T : 終端記号の集合
 - P : 書換規則の集合
 - σ : 開始記号 ($\sigma \in V_N$)
- P は次の形
 - $\alpha \rightarrow \beta$ $\alpha \in (V_N \cup V_T)^+, \beta \in (V_N \cup V_T)^*$
- 開始記号 σ から書換規則を順次適用して生成される言語 $L(G)$ を0型言語と呼ぶ



チヨムスキーの階層

- 0型文法 (チューリング機械)
 - もっとも制約のない形式文法
- 1型文法 (文脈依存文法)
 - $\alpha \rightarrow \beta$ ($|\alpha| \leq |\beta|$)
- 2型文法 (文脈自由文法)
 - $A \rightarrow \beta$ ($A \in V_N, \beta \in (V_N \cup V_T)^+$)
- 3型文法 (正規文法)
 - $A \rightarrow aB$
 - $A \rightarrow a$
 - $(A, B \in V_N, a \in V_T)$



文脈自由文法 (1/3)

(Context Free Grammar, CFG)

- $G (= \langle V_N, V_T, P, \sigma \rangle)$
 - V_N : 非終端記号の集合
 - V_T : 終端記号の集合
 - P : 書換規則の集合
 - σ : 開始記号 ($\sigma \in V_N$)
- P は次の形
 - $A \rightarrow \alpha$ ($A \in V_N, \alpha \in (V_N \cup V_T)^+$)



文脈自由文法 (2/3)

(Context Free Grammar, CFG)

- 直接導出

- $A \rightarrow \alpha$ の規則があれば、 $\beta A \gamma$ は $\beta \alpha \gamma$ に書き換えられる ($\alpha, \beta, \gamma \in (V_N \cup V_T)^*$, $A \in V_N$)

$$\beta A \gamma \Rightarrow \beta \alpha \gamma$$

- 導出

- $\alpha_1 \Rightarrow \alpha_2, \alpha_2 \Rightarrow \alpha_3, \dots, \alpha_{n-1} \Rightarrow \alpha_n$ であるとき、

$$\alpha_1 \xRightarrow{*} \alpha_n$$



文脈自由文法 (3/3)

(Context Free Grammar, CFG)

- 文形式
 - 記号列 $\alpha \in (V_N \cup V_T)^*$ が開始記号 σ から導出されるとき ($\sigma \xRightarrow{*} \alpha$)、 α を文形式と呼ぶ
- 文
 - 終端記号のみからなる文形式
- 言語 $L(G)$
 - 文法 G から導出される文全体の集合

$$L(G) = \{s \in V_T^* \mid \sigma \xRightarrow{*} s\}$$



CFGの例

- $G (= \langle V_N, V_T, P, \sigma \rangle)$
 - $V_N = \{S, NP, NP1, SUBJ, OBJ1, V, VP1\}$
 - $V_T = \{\text{香織, 恵, 電子メール, プレゼント, と, が, を, 送った, 読んだ}\}$
 - $P = \{S \rightarrow SUBJ VP1, S \rightarrow SUBJ V, SUBJ \rightarrow NP \text{ が, } VP1 \rightarrow OBJ1 V, OBJ1 \rightarrow NP \text{ を, } NP \rightarrow S NP, V \rightarrow \text{送った, } V \rightarrow \text{読んだ, } NP \rightarrow \text{香織, } NP \rightarrow \text{恵, } NP \rightarrow \text{電子メール, } NP \rightarrow \text{プレゼント, } NP \rightarrow \text{香織 } NP1, NP \rightarrow \text{恵 } NP1, NP1 \rightarrow \text{と } NP\}$
 - $\sigma = S$



CFGの限界

表現できる言語の限界

- $L(G) = \{uv^iwx^iy \mid i = 0, 1, 2, \dots\}$ となるG
 - 正規文法では表現できない
 - CFGは表現できる
- $L(G) = \{a^n b^n c^n \mid n = 0, 1, 2, \dots\}$ となるG
 - CFGは表現できない
 - 文脈依存文法は表現できる



C F Gの限界

- 文法機能の解析不足

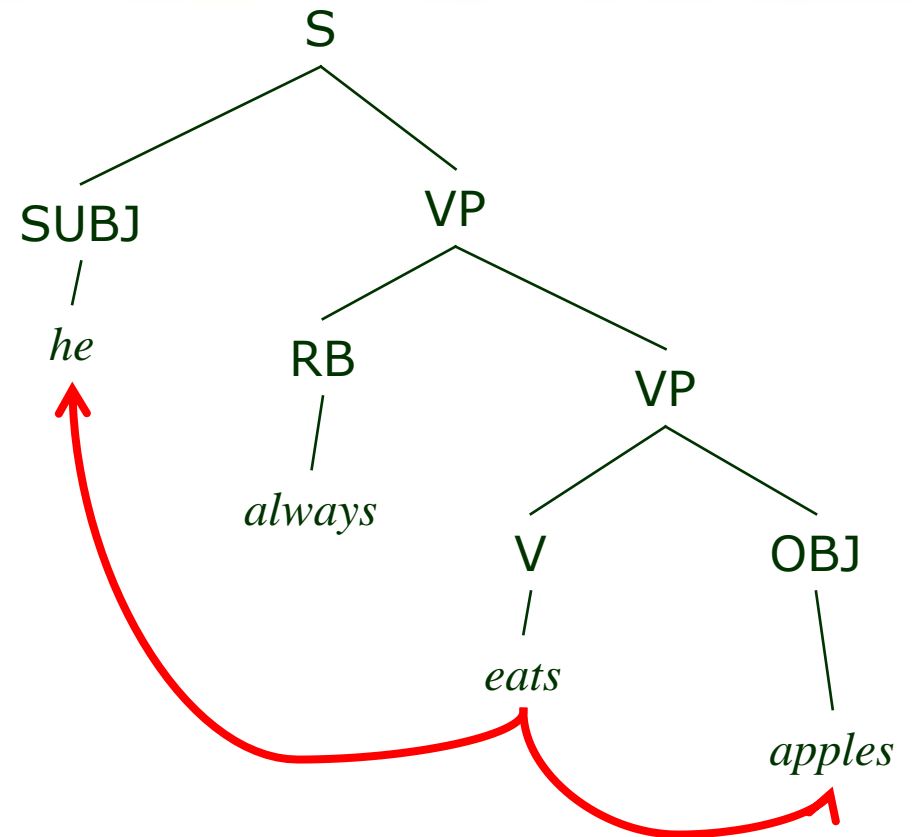
S → SUBJ VP

VP → RB VP

VP → V OBJ

VP → V

NP → NP which S



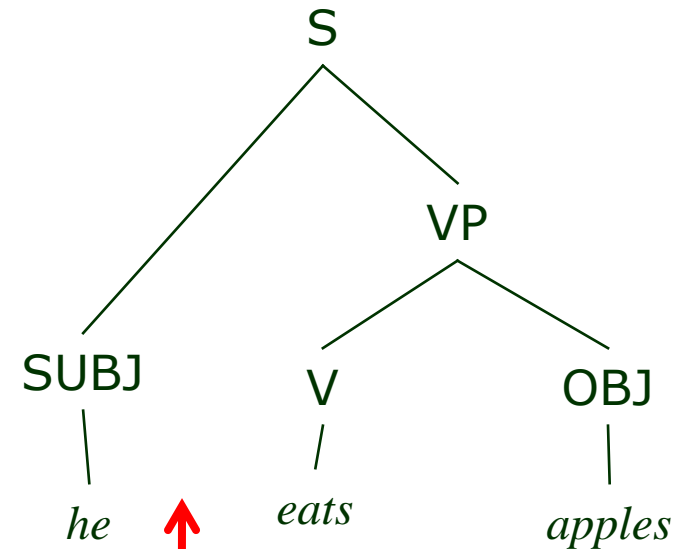
eats の主語、目的語が何かわからない



C F Gの限界

- 文法機能の解析不足

$S \rightarrow \text{SUBJ } V \text{ OBJ}$
というルールにすれば
直接の関係はわかるけど、
助動詞などが間にはいると、
 $S \rightarrow \text{SUBJ always } V \text{ OBJ}$
と助動詞の数だけ規則が増え
ていってしまう



always は？



C F Gの限界

- 長距離依存

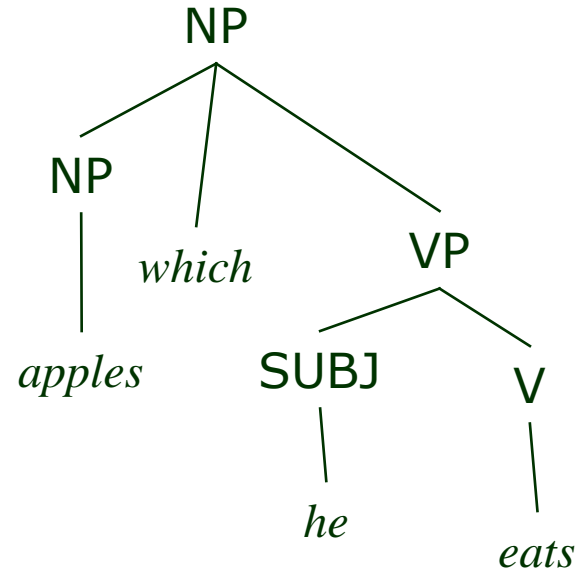
S → SUBJ VP

VP → RB VP

VP → V OBJ

VP → V

NP → NP which S



eats の目的語がapplesということがわからない

CFGの構文解析

CFG
S → SUBJ VP
SUBJ → N が
N → N と N
VP → PP VP
PP → N と
VP → OBJ V
OBJ → N を
N → 太郎
N → 花子
N → 映画
V → 褒める

- ある文sが与えられた時、その可能な構文木を全て求める
- ある文sが与えられた時、可能な構文木の中で最も正しそうな構文木を求める

”太郎が花子と映画を褒める”



CFGの構文解析

CFG

S → SUBJ VP

S → VP

SUBJ → N が

VP → OBJ V

VP → V

OBJ → N を

N → S N

N → N と N

V → 送った

V → 読んだ

N → 香織

N → 恵

N → 電子メール

“香織 が 恵 が 送った 電子メール を 読んだ”

“香織 と 恵 が 送った 電子メール を 読んだ”

言語学での発展



言語学

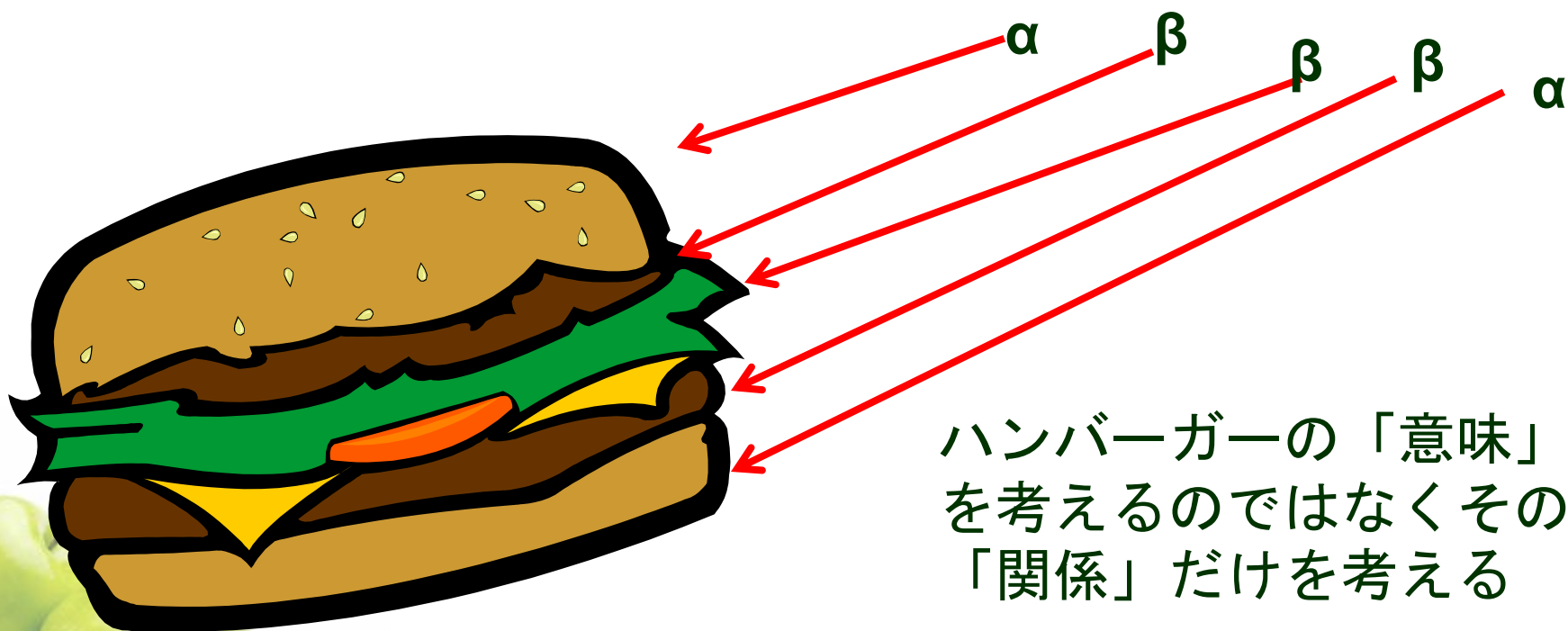
- ソシユールの構造主義
 - 言語をばらばらな要素の羅列と考えず、要素の関係による構造または体系として捉える
 - 個々の要素をより大きな構造全体の中で位置づける
 - 各要素の具体的な特徴ではなく、他の要素の相対的な関係（差異）から規定される
 - 特定の言語の時間にそった変化（通時態）を追うのではなく、特定の言語の一時期の状態（共時態）を追う



構造主義

- 誤解をおそれずにたとえると、、、

ハンバーガーとは、「 $\alpha \beta * \alpha$ 」である



構造主義といえは、、、

- レヴィストロース

- 「『構造』とは要素と要素間の関係とからなる全体であって、この関係は一連の変形過程を通じて不変の特性を保持する」

- 顔

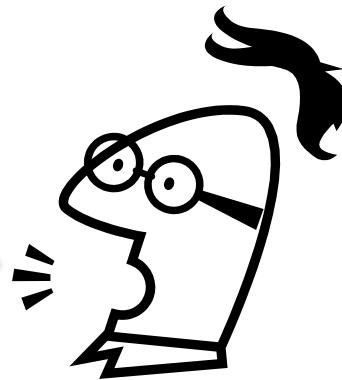
- 目、口、鼻、耳などのパーツからなるが、個々のパーツの関係で顔が構成される
- 世の中にある多種多様な顔は「顔」という関係の変形



つまり、構造主義とは、、、

- 「アリストテレスが死ぬ」とはどういうことですか？
 - 死ぬ(アリストテレス)
 - 同義語(死ぬ, 亡くなる), 同義語(死ぬ, 死亡する),
.....

これだけ?!



構造言語学

- 構造言語学
 - ブルームフィールド以後の言語学
 - 発話の資料体（コーパス）を構成する要素の帰納的・分類的研究
 - コーパスを構成する要素
 - 音素
 - 形態素



チヨムスキーの変形生成文法

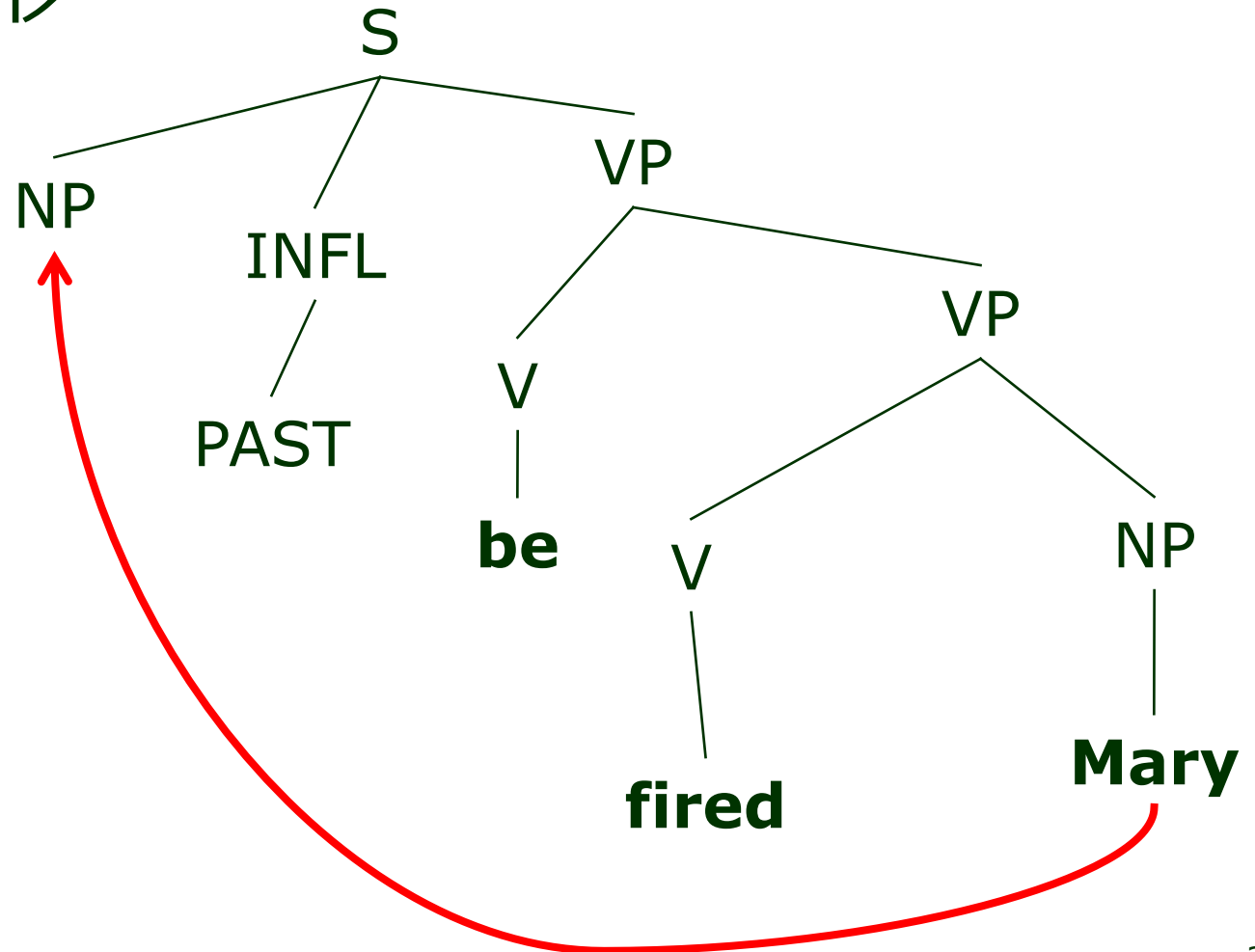
- 構造言語学への批判
 - 帰納的・分類的から演繹的・説明的
 - コーパスに直接みえる構造ではなく、人間の言語を生成する能力に注目
 - 言語の構造や体系を重視・優先
 - CFGによる文法の記述
- CFGへの批判
 - 自然言語を記述するには不足
 - CFG + 変形操作の文法 (= 変形生成文法)



チヨムスキーの変形生成文法

- CFG + 変形

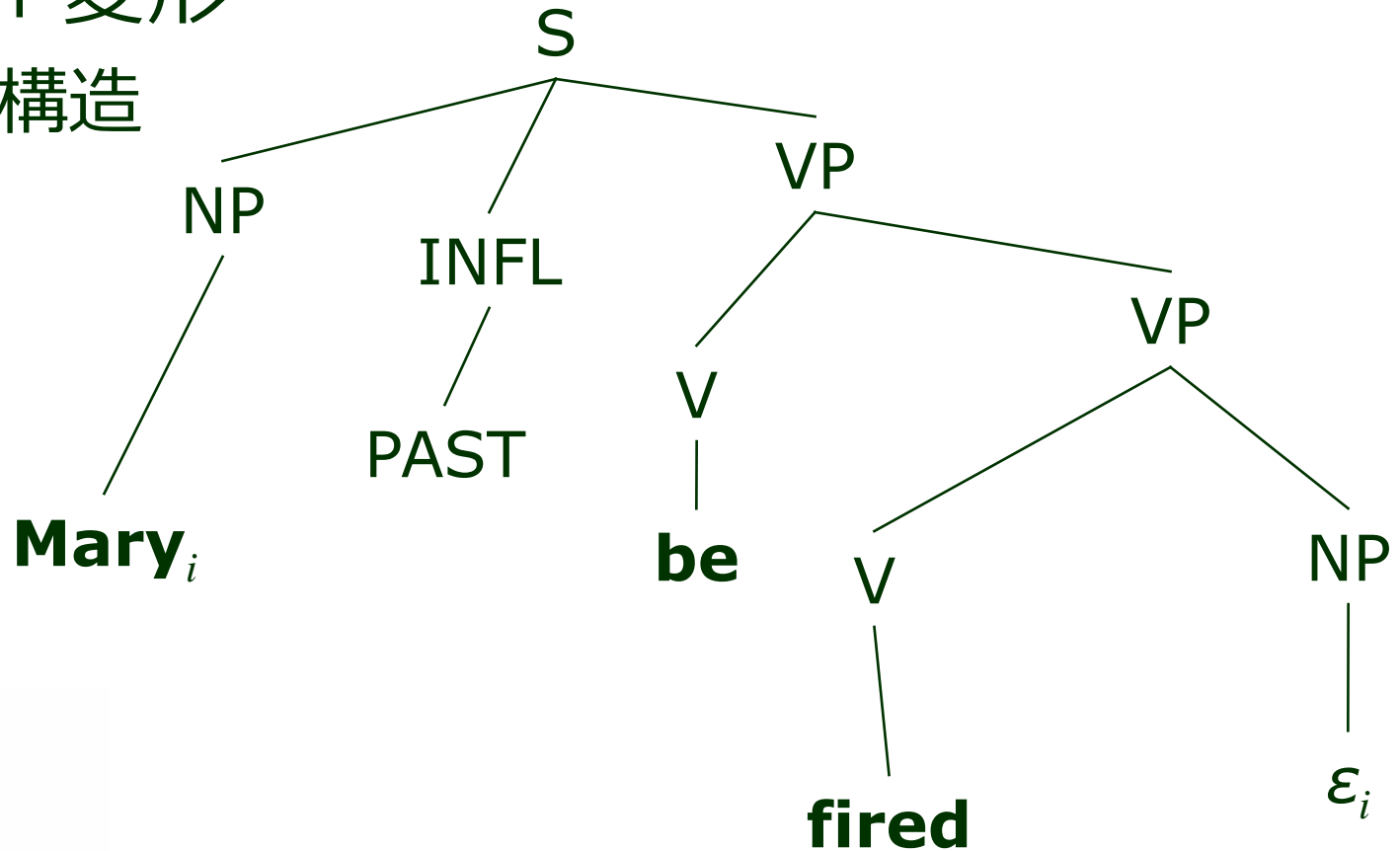
- d-構造



チヨムスキーの変形生成文法

- CFG + 変形

- s-構造



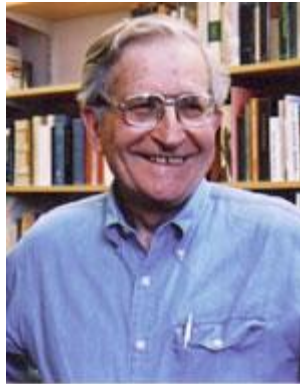
チヨムスキーの変形生成文法

- 普遍文法 (Universal Grammar)
 - 個々の言語は普遍文法のパラメータの設定により得られる
- Xバー理論
 - 名詞が形容詞や冠詞と結びつき名詞句となる
- θ 理論
- 下位範疇化
- 適格文/非文
 - 適格文、非文の区別により、正しい文法を模索する \Leftrightarrow Meaning-Text Theory

チヨムスキーという人

- チヨムスキー ???

構造主義的



認知科学的

政治的にも有名？



チョムスキー以後

- 様々な文法理論
 - GPSG
 - HPSG
 - LFG
 - TAG
 -

しかし、基本的な考え方はチョムスキーと同じ



まとめ

- CFG
- 言語学での発展

- 連絡・資料

<http://aiweb.cs.ehime-u.ac.jp/~ninomiya/ai2/>

