



# HPSG文法開発(前半)

二宮 崇

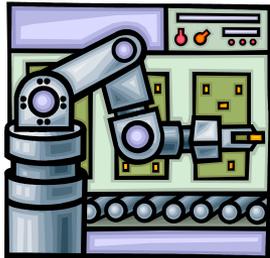
# 今日の講義の予定

- 文法開発（前半）
  - 合理主義的文法
  - 経験主義的文法



# いわゆる“自然言語処理”

“太郎は花子が好きだ”

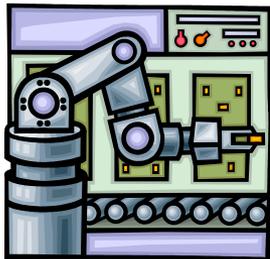
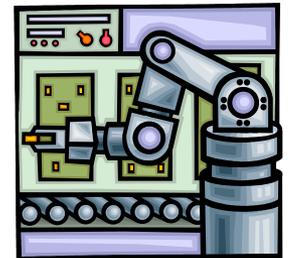


形態素解析

太郎/名詞 は/助詞 花子/名詞 が/助詞 好きだ/形容動詞

(文 (名詞句-主語太郎/名詞 は/助詞) (名詞句-目的語  
花子/名詞 が/助詞) (動詞句 好きだ/形容動詞))

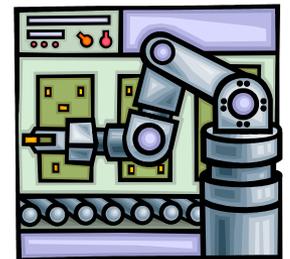
構文解析



意味解析

太郎/名詞/主語/動作主/人物  
花子/名詞/目的語/対象/人物  
好きだ/動詞/動作主-太郎/対象-花子

文脈解析

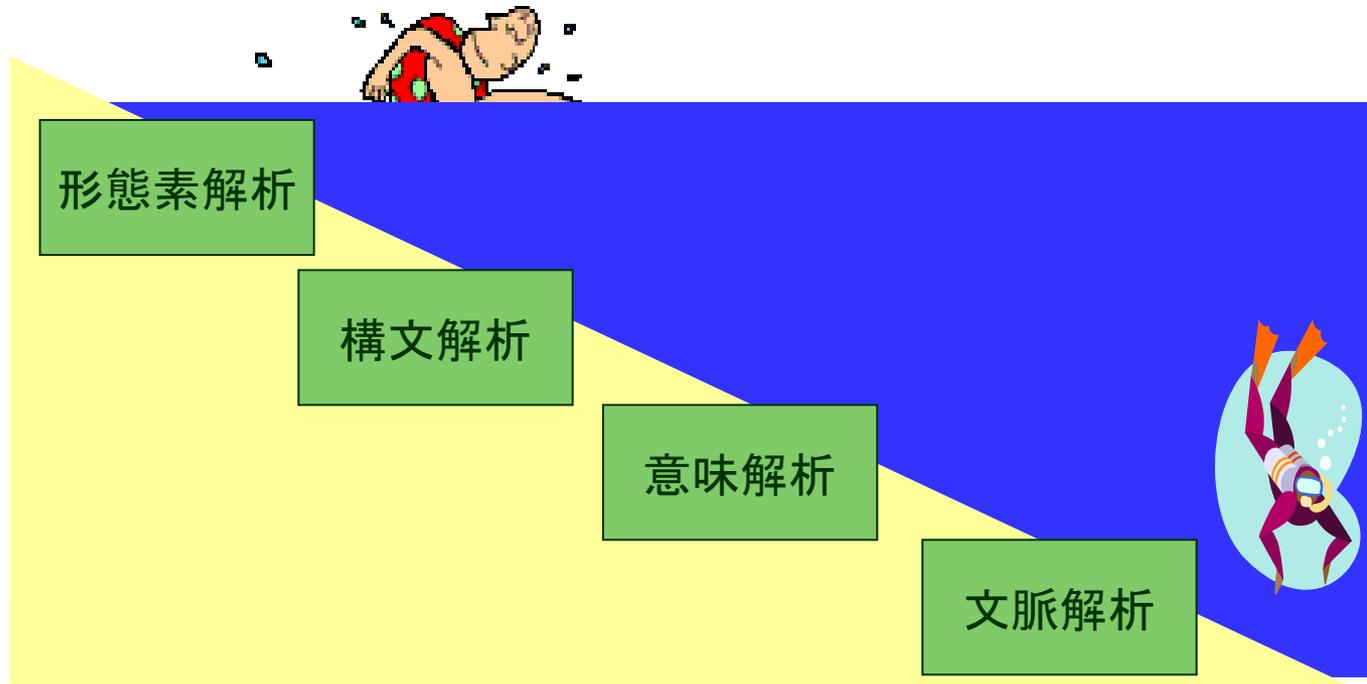


...



# “浅いところ”から“深いところ” へ

理想



# “浅いところ”から“深いところ”



現実

形態素解析



思いの他  
深い！

TAG, LFG, HPSGなどあまたの複雑精巧な文法が提案・研究されてきたにも関わらず、実テキストを解析できる文法はなかなかできなかった

構文解析

意味解析

文脈解析



# “深海”を目指すよりも“浅瀬”を



形態素解析

構文解析

意味解析

文脈解析

- ・ コーパスベース
- ・ 統計モデル
- ・ 機械学習

複雑精巧な文法理論  
に頼らなくてもそこ  
そこの出力が得られ  
る

# 文法開発の難しさ

- さて、いったい何が難しくて文法開発がうまくいかなかったのだろうか？



# 構造と言語能力と文法理論

## 文法

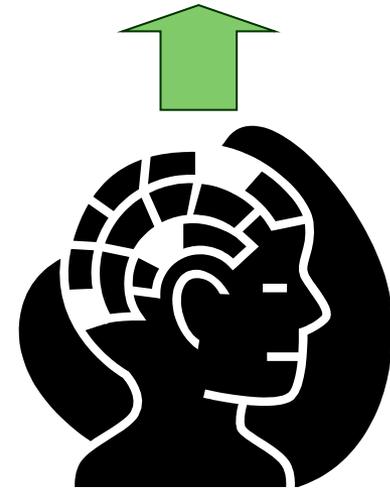


辞書

$S \rightarrow NP VP$   
 $NP \rightarrow DET N$   
 $NP \rightarrow N$   
...

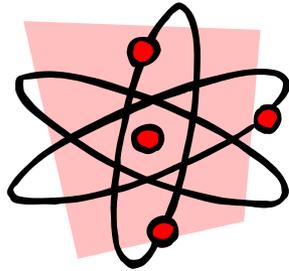
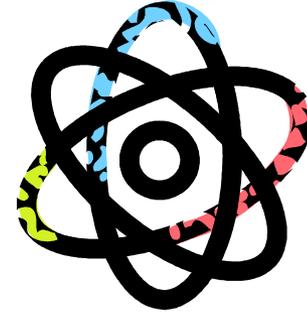
文法規則  
(=生成規則+制約)

適格文、非文を人間に判断させることによって、人間がもつ言語能力の規則性 (=文法) を発見する



# 自然科学と文法理論

不可知な真の自然



原子、分子、  
クォーク

理論化、検証を繰り返す  
ことによって、真の自然  
の姿により近づく



# 自然科学と文法理論

不可知な真の文法

S → NP VP  
NP → DET N  
NP → N  
...



S → NP VP  
NP → DET N  
NP → N  
...



文法規則、辞書、シソーラス

理論化、検証を繰り返す  
ことによって、真の文法  
の姿により近づく



# 文法理論と科学的サイクル

思考実験

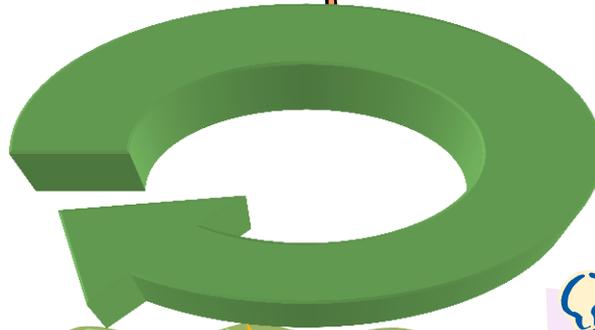
コーパスに対する検証

理論の検証



データ収集・観察・分析

コーパス収集  
コーパス開発  
コーパス分析



カテゴリー化  
文法理論  
辞書項目

理論化



# どこに落とし穴があったのか？



# アウトライン

- 導入
- 合理主義的文法
- 経験主義的文法
- 文法開発の再解釈と展望
- 合理主義的文法と経験主義的文法を超えて





# 合理主義的文法

# 合理主義的文法

- ・ 文法を人間が定義、分類、記述する
- ・ 辞書と文法規則を開発
- ・ コーパスは検証のための副次的存在

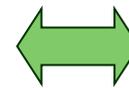
S → NP VP  
NP → DET N  
NP → N  
...



文法規則、辞書



検証

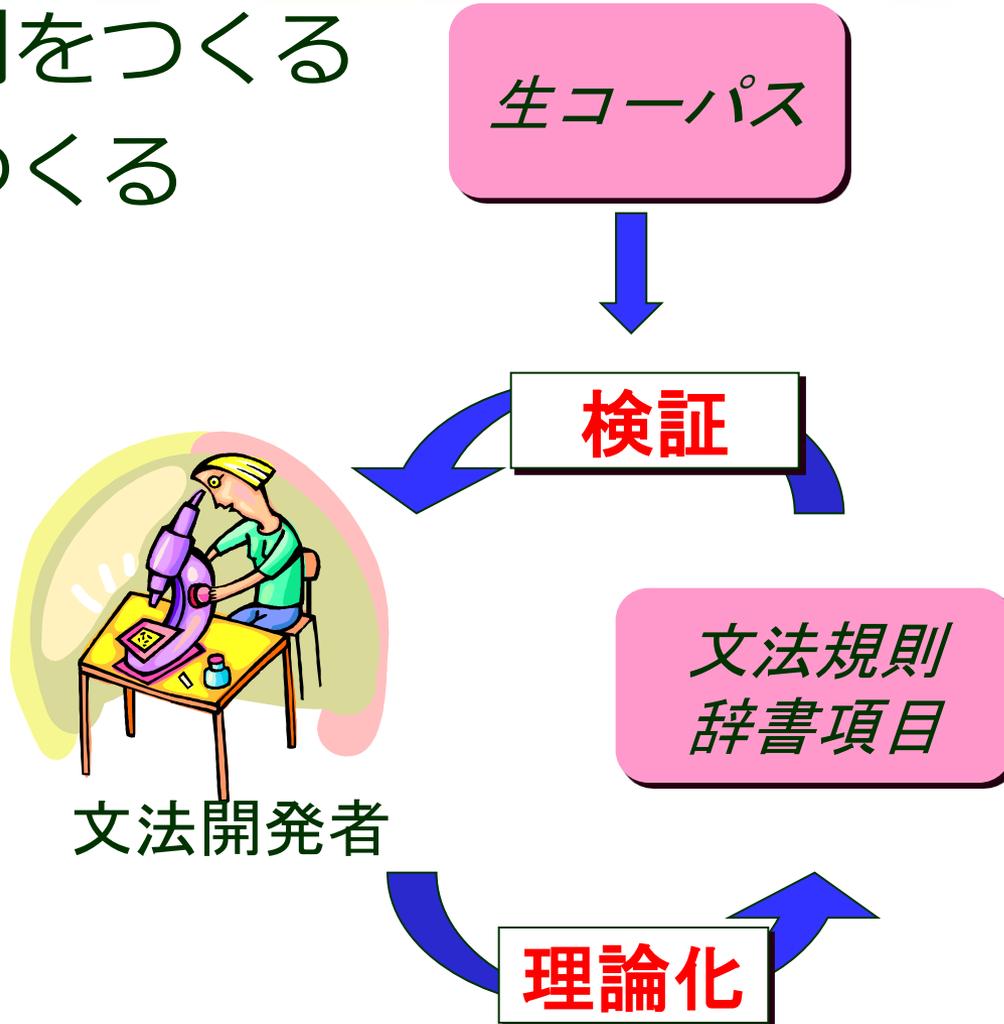


コーパス



# 合理主義的文法の文法開発

- 文法規則をつくる
- 辞書をつくる



# 有名な合理主義的文法

- Core Language Engine (English) [Alshawi 1992]
- TAG [Joshi et al. 1996]
  - XTAG (English/Korean) [XTAG Research Group 1995]  
<http://www.cis.upenn.edu/~xtag/>
  - FTAG (French) [Abeillé et al. 2000]
- LFG [Bresnan 1982]  
<http://www.essex.ac.uk/linguistics/LFG/>
  - ParGram (English, Chinese, French, German, Norwegian, Japanese, Turkish, Urdu, Welsh, Malagasy, Arabic, Hungarian, Vietnamese) [Butt et al. 2002] <http://www2.parc.com/isl/groups/nltt/pargram/>
  - English XLE [Riezler et al. 2002; Kaplan et al. 2004]
  - German XLE [Forst and Rohrer 2006]
  - Japanese XLE [Masuichi and Okuma 2003]



# 有名な合理主義的文法

- HPSG [Pollard et al. 1994]
  - DELPHIN (English, Japanese, German, Spanish, Norwegian, Modern Greek, Korean, Italian) [Bender et al. 2002]
    - LinGO ERG (English) [Flickinger 2002]
    - JACY (Japanese) [Melanie et al. 2002]
  - Babel (German) [Stefan Müller 1996]
  - ALPINO (Dutch) [Bouma et al. 2002]
- RASP (English) [Carroll and Briscoe 2002]



# HPSG

- 現代の言語学において代表的な文法理論
- 文法開発や高速化の研究もさかん
- 中心的概念：文法=辞書項目+文法規則
  - 辞書項目：単語固有の構文・意味的性質を記述する
  - 文法規則：構文木の一般的規則性を規定する



# 合理主義的文法開発の現状と問題点

- 大規模かつ複雑な構造を実装するのは非常に難しい
- 複雑な文法を効率的に開発するために、様々な文法開発ツールが開発された
  - XTAG [XTAG Research Group 1995]
  - ConTroll [Götz et al. 1997]
  - LKB [Copestake et al. 1999]
  - [incr tsdb()] [Oepen et al. 2000]
  - XLE [Butt et al. 2002]
- しかし、実世界のテキストを網羅的に解析できる文法の開発は難しかった [Baldwin et al. 2004]
  - LinGO ERGでBNCコーパスを解析
    - 辞書登録されている単語で構成されている文の割合
      - 32%
    - 構文解析結果を一つ以上出力する割合
      - $32\% \times 57\% = 18\%$  (未知語対策をしても43%の文は解析できない!)
    - 構文解析結果の中に正解が含まれている割合
      - $32\% \times 57\% \times 87\% = 16\%$



# 合理主義的文法開発の最先端

- Grammar Matrix (in DELPHIN Project) [Bender et al. 2002]
    - 多言語文法開発のための文法コンポーネントの共通化 (English, Japanese, German, Spanish, Norwegian, Modern Greek, Korean, Italian)
  - Optimality Theory (in LFG XLE) [Frank et al. 1998]
    - 曖昧性解消のために規則に優先順位を付与
  - ツリーバンク開発 (後述)
    - 曖昧性解消のための学習用
    - 評価用
- c.f. The PARC 700 Dependency Bank [King et al. 2003]  
LinGO Redwoods [Oepen et al. 2002]  
Hinoki [Bond et al. 2004]



# 合理主義的文法開発の前提

## ● 網羅性

- 実世界の文を解析するためには、あらゆる文を網羅する  
大規模な文法が必要
- 継続的に文法を修正・拡張

## ● 識別性

- 適格文のみ構文木が導出できて、非文は導出されない
- 文法的に解釈できる構文木はすべて出力
  - どの構文木をもっともらしい解とするかは、選好(preference)の問題であって、別モジュールで解決すべき問題
- 一文に対し文法的に解釈できる解の数は少ないほうがよ

い



# 合理主義的文法開発の問題点(1/2)

- 網羅性と一貫性のトレードオフ
  - 文法を修正・拡張する際、一貫性・無矛盾性を保つのが非常に困難
  - コーパス中の問題の一箇所を改良すると、他の箇所に悪影響が及ぶ
  - ポリシーの変更により10万オーダーの辞書項目を大幅に書き換える必要がある場合もある
  - 文法修正により改良されたのか改悪されたのか明確に判断できない



# 合理主義的文法開発の問題点 (2/2)

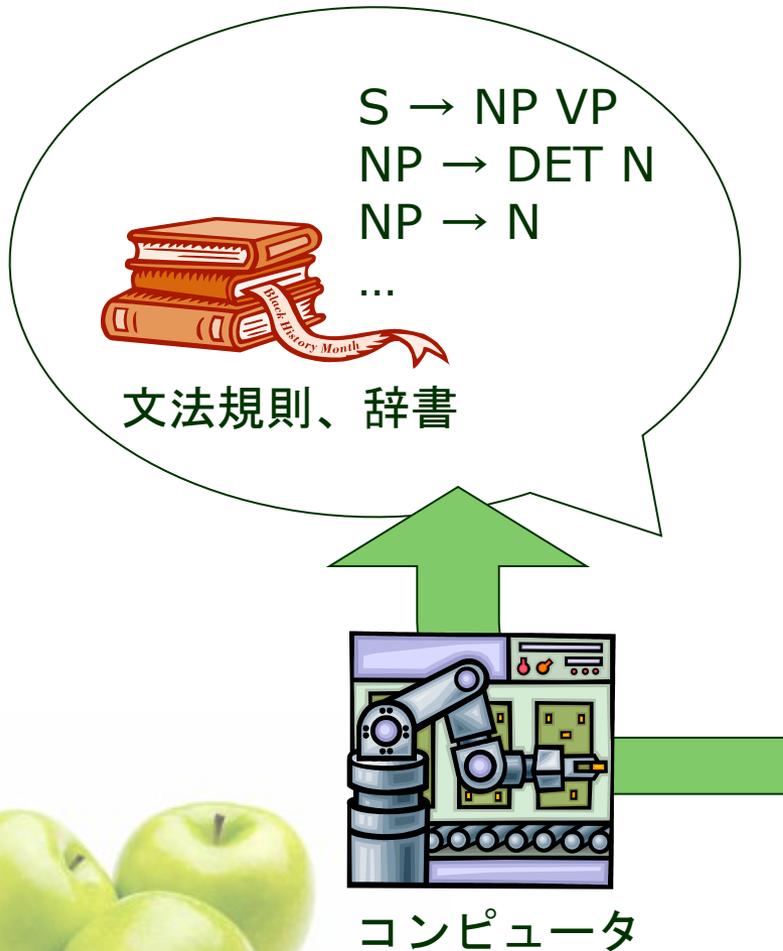
- 曖昧性解消の必要性
  - 実際のアプリケーションは一つの文に対し一つの解析結果を要求する
  - 合理主義的文法の文法開発では曖昧性解消の問題を先送りしている



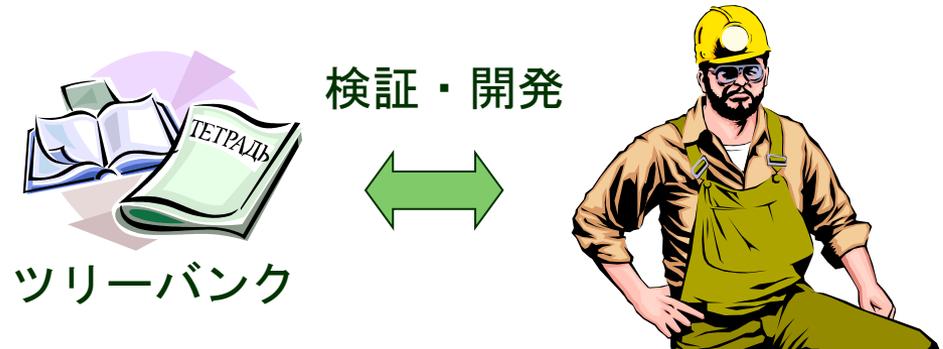


# 經驗主義的文法

# 経験主義的文法

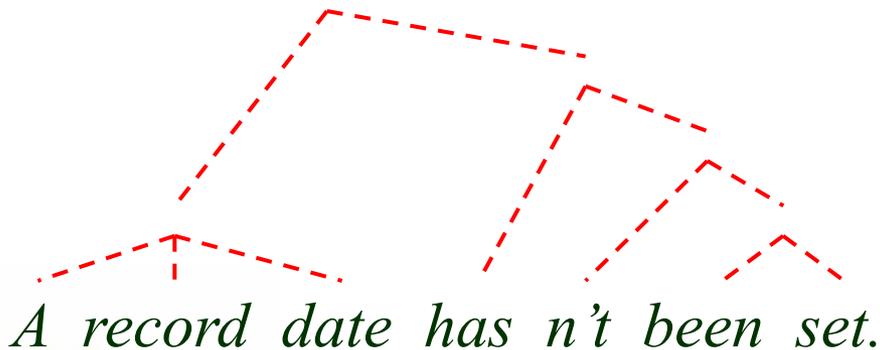


- 人間が文法を直接定義するのは困難
- 構文木の实例 (ツリーバンク) に基づく定量的評価が必要
- 文法はツリーバンクから導出  
(ツリーバンク文法)



# ツリーバンク

- 実世界の文に対して人手で構文木を付与する
- 明示的な文法を仮定しない
- 構造は開発者の言語直感とガイドラインに依存
  - ガイドラインはあるが、文法で定義されるような「何が正解か」の客観的基準は存在しない



*A record date has n't been set.*



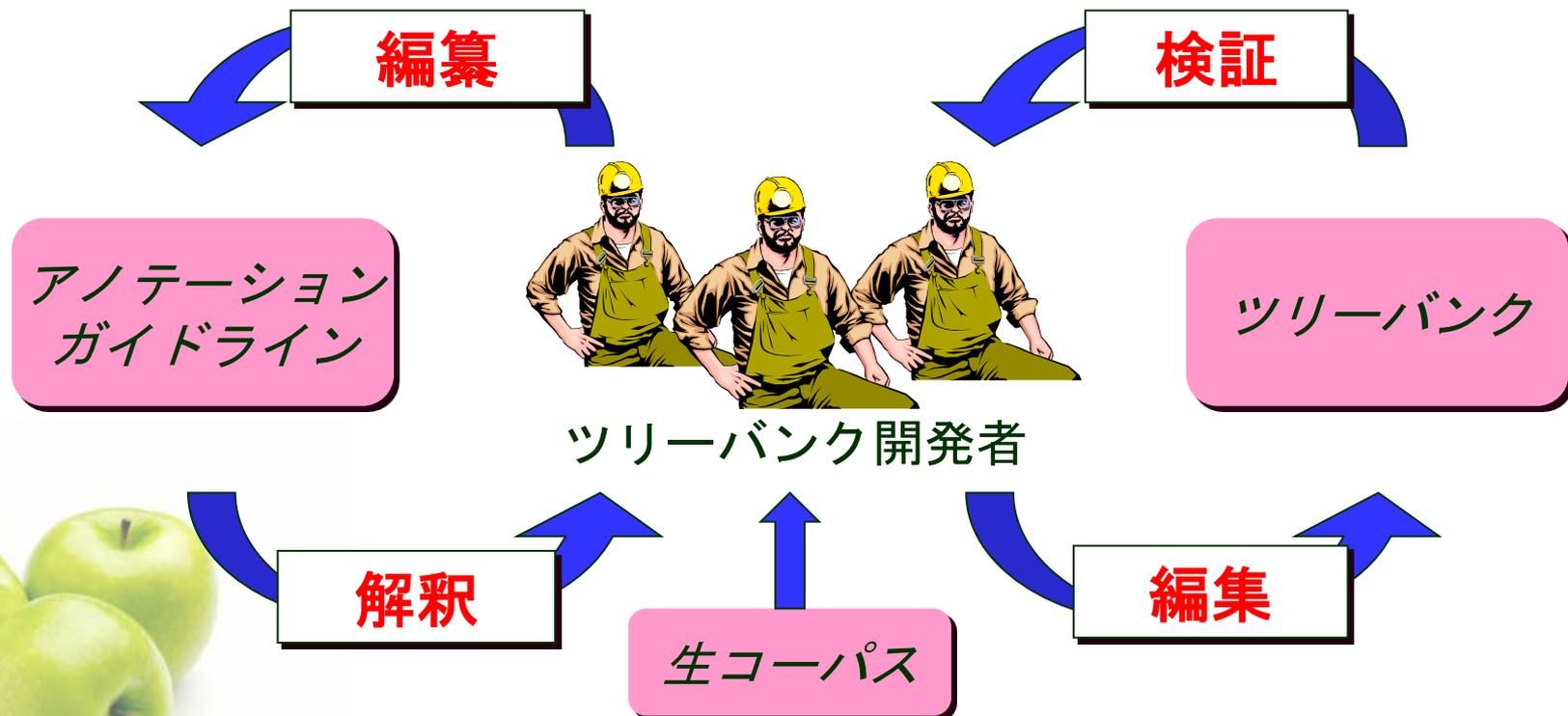
# 有名なツリーバンク

- 構文木や係り受け木を人手で付与したコーパス（ツリーバンク）の登場
  - Penn Treebank [Marcus et al. 1993]
  - SUSANNE [Sampson 1995]
  - TIGER Treebank [Brants et al. 2002]
  - Prague Dependency Treebank [Hajic 1998]
  - Verbmobil [Hinrichs et al. 2000]
  - EDRコーパス [EDR 1995]
  - 京都大学テキストコーパス [黒橋ら 1997]
  - 日本語話し言葉コーパス [前川ら 2000]



# ツリーバンクの開発過程

- ツリーバンク開発者（アノテータ）による手作業
- アノテータのためのマニュアル（アノテーションガイドライン）による品質管理



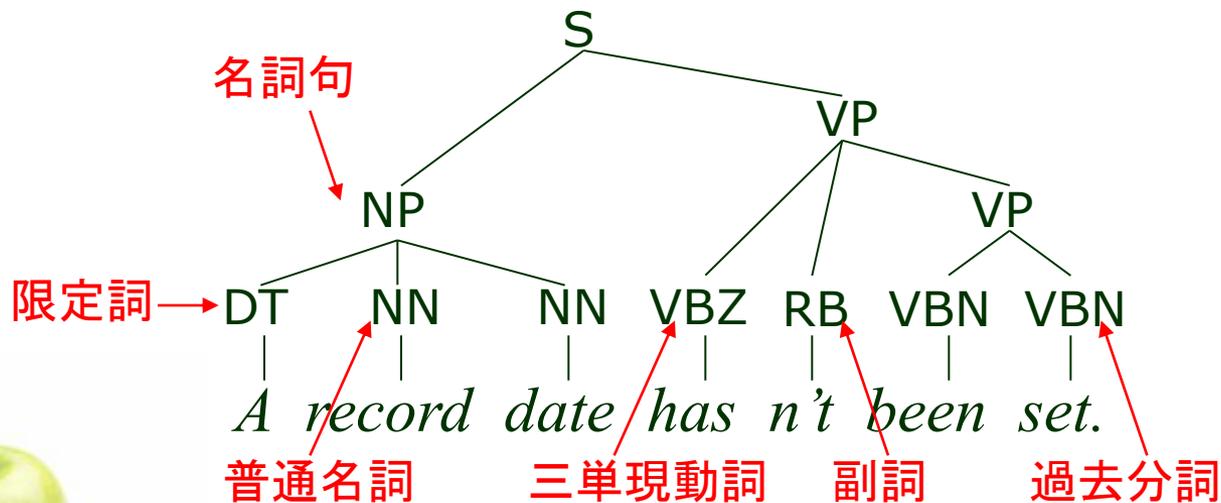
# Penn Treebank (1/2)

- 構文木が付与された最初の大規模英語ツリーバンク [Marcus et al. 1993]
- 様々な分野の英語テキストを収録
  - Wall Street Journal (新聞) 約5万文、100万語
  - ATIS (航空券予約の会話)
  - Brown (様々な分野のテキスト)
  - Switchboard (電話の自由発話)



# Penn Treebank (2/2)

- 品詞：NN（普通名詞），VBZ（三単現動詞） ...
- 構文木：NP（名詞句），VP（動詞句） ...
- Function tag, null element: 述語項構造を計算するための付加情報（詳細省略）



# Penn Treebank

## アノテーションガイドライン

### “Bracketing Guidelines for Treebank II Style Penn Treebank Project” Bies et al. 1995

- 1. An Overview of Basic Clause Structure
- 2. Notation
- 3. Punctuation
- 4. Null Elements
- 5. Pseudo-Attach
- 6. Copular Verbs
- 7. Coordination
- 8. Shared Complements and Modifiers in Coordinated Structures
- 9. WH-Phrases
- 10. Subordinate Clauses
- 11. Modification of N
- 12. Titles
- 13. Gerunds and Participles
- 14. Infinitives
- 15. Small Clauses and their near relatives
- 16. Clefts
- 17. It-Extraposition
- 18. Subject-Raising Predicates
- 19. Whether *it* and Referential *it*
- 20. Existential *there*
- 21. *Tough*-Clefts
- 22. Comparatives
- 23. “Financialspeak” conventions
- 24. Numbered Lists
- Correlative *the*-Clauses
- Orphans

全318ページ!

## アノテーションガイドラインの例 (1/3)

- 基本的には自然言語による解説とたくさんの例示
- 1.1.4 名詞句内の補語
  - 名詞にかかるPPがadjunctなのかargumentかを区別するのは難しいので、たんにNPにくっつける

(NP (NP a teacher)

(PP of

(NP chemistry)))

- ただし、補文がかかる場合は次のようにする

(NP the belief

(SBAR that

(S the world is flat)))



# Penn Treebank

## アノテーションガイドライン (2/3)

- 13.3.5 ADJP vs. S
  - 動名詞は2種類の解釈がある: 形容詞的名詞句修飾 (ADJP) vs 動名詞句 (S)
  - “Flying planes can be dangerous”
    - (a) (S (NP-SBJ Flying planes)  
(VP can  
(VP be  
(ADJP-PRD dangerous))))
    - (b) (S (S-NOM-SBJ (NP-SBJ \*)  
(VP Flying  
(NP planes)))  
(VP can  
(VP be  
(ADJP-PRD dangerous))))
  - 判断がつかないときのデフォルトは(a)



# Penn Treebank

## アノテーションガイドライン (3/3)

- Small clause

- 15.3.1. to-不定詞に関する句のアノテーション

- 1. monotransitive (S) vs. ditransitive (NP+S)

(a) (S (NP-SBJ This)

(VP does not

(VP allow

(S (NP-SBJ the myste

(VP to

(VP invade (N

(b) (S (NP-SBJ He)

(VP told

(NP-1 me)

(S (NP-SBJ \*-1)

(VP to

(VP wake (NP you))))))

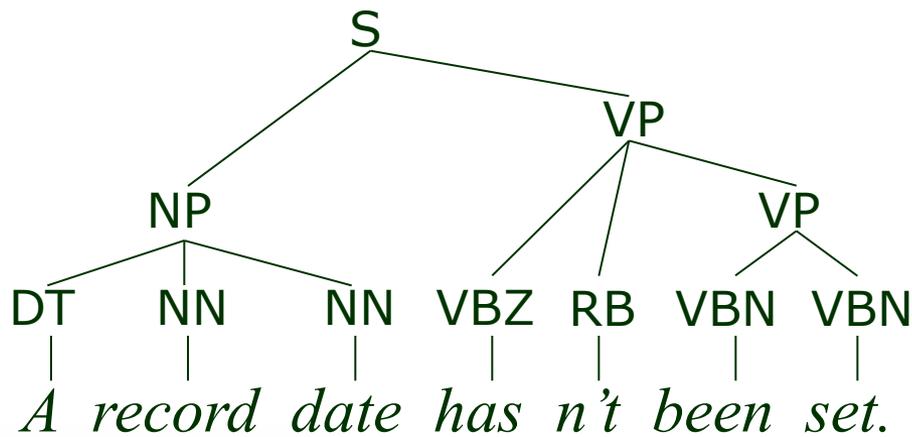
*advise, ask, beg, beseech, challenge, command, counsel, detail, direct, enjoin, exhort, forbid, implore, incite, inform, instruct, invite, order, persuade, pray, promise, remind, request, recommend, teach, tell, urge*

の場合は**(b)**と解釈



# ツリーバンクから文法を抽出する

- ツリーバンクの背後にある文法を自動抽出
  - 潜在的な規則性を自動獲得できるはず



文法抽出

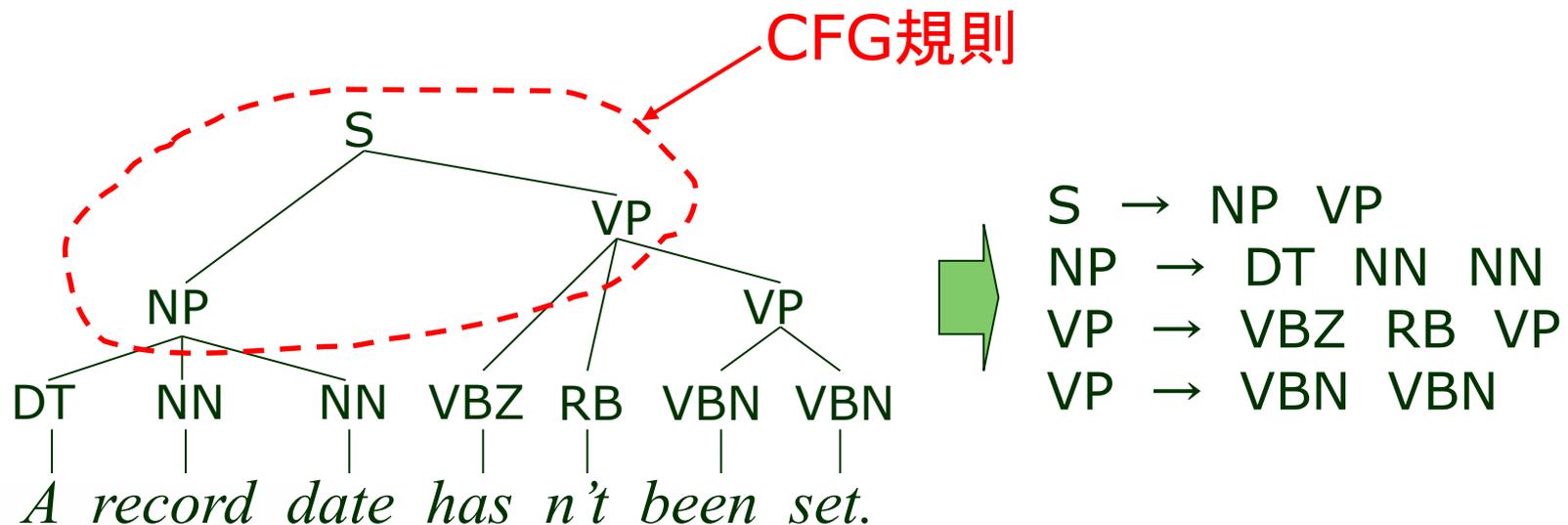


ツリーバンク  
開発



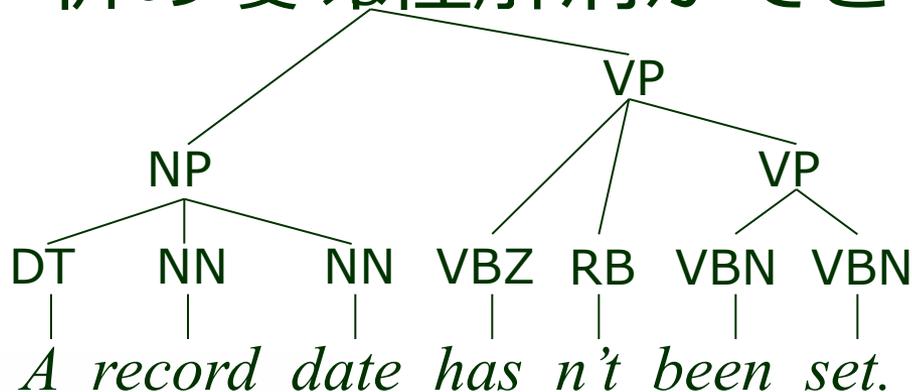
# 確率CFGの自動抽出(1/2)

- ツリーバンクの各分岐をCFG規則だと仮定して抽出する [Charniak 1996; 1997] c.f. [Sekine1995]



# 確率CFGの自動抽出(2/2)

- ツリーバンクでの出現頻度から確率値を推定
- 確率値最大の木を探索することで、構文解析の曖昧性解消ができる

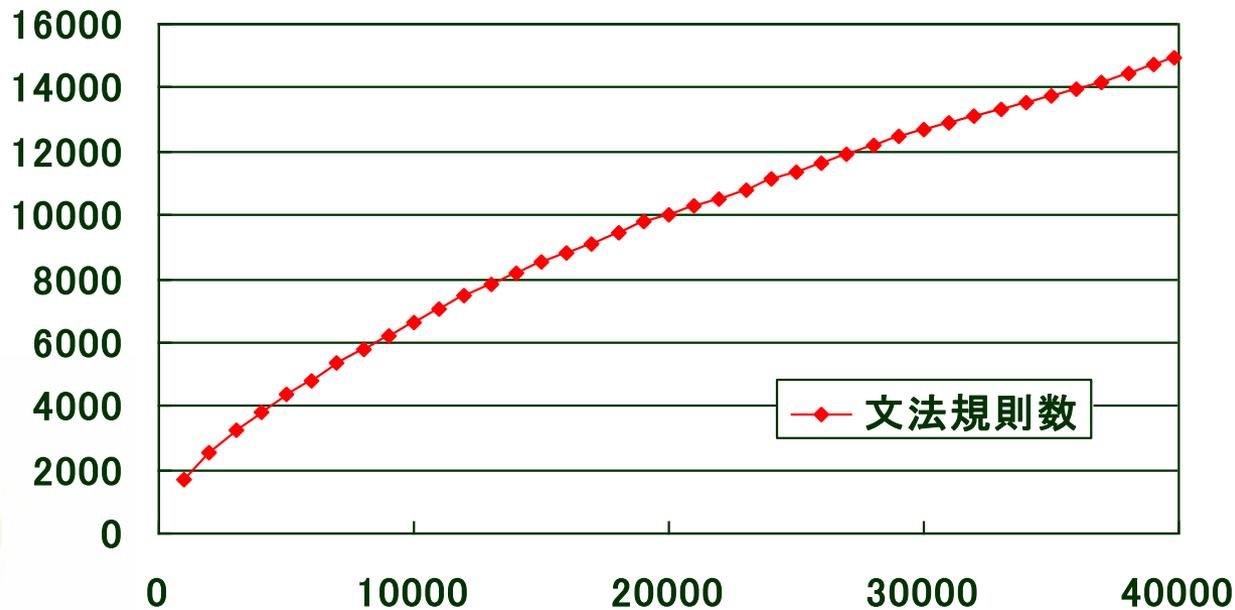


S	→	NP	VP	0.5	
NP	→	DT	NN	NN	0.03
VP	→	VBZ	RB	VP	0.02
VP	→	VBN	VBN	0.1	



# 問題点（１）：文法が大きい

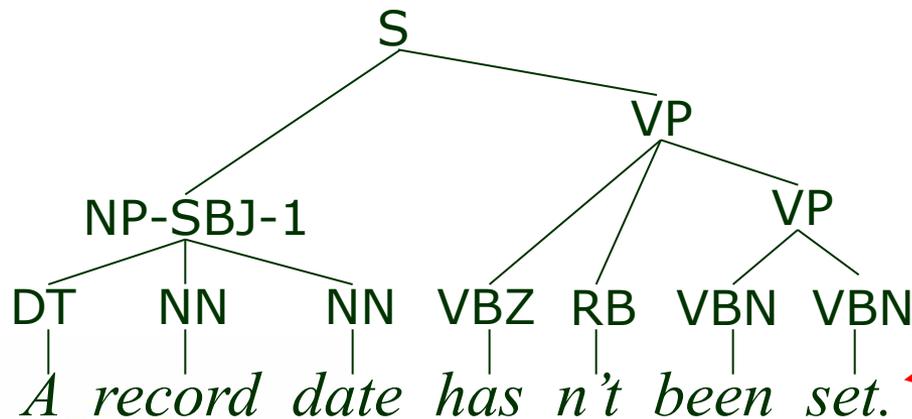
- 40,000文から約15,000のCFG規則
  - CFG規則数が収束しない [Carpenter et al. 1997]
- 抽象化・一般化しきれていない





# 問題点（3）：構造が浅い

- CFG構文木しか出力できない
- 意味構造へのマッピングがない
  - 有用な情報が得られない
  - 文生成に使えない



主語、目的語はどこ？  
時制、アスペクトは？



# ツリーバンク文法の改良

## (1) 文法が大きい

- CFG規則の自動圧縮 [Krotov et al. 1998; 1999]
- **CFG規則の確率モデル化** [Magerman 1995; Collins 1997; Charniak 2000]

## (2) 精度が低い

- **非終端記号の細分化** [Magerman 1995; Collins 1996; 1997; Johnson 1998; Charniak 2000]

## (3) 構造が浅い → 後述



# CFG規則の確率モデル化

- Markov Grammar: CFG規則を確率的に生成する [Collins 1997; Charniak 2000]

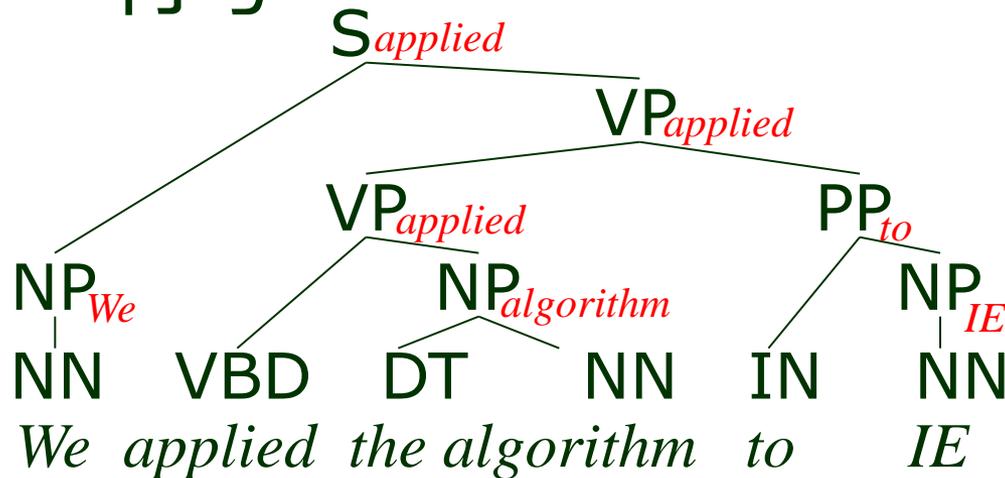
$$\begin{aligned} & p(\text{NP} \rightarrow \text{DT NN NN} \mid \text{NP}) \\ &= p(\text{NN} \mid \text{NP}) p(\text{NN} \mid \text{NN, NP}) p(\text{DT} \mid \text{NN, NN, NP}) \end{aligned}$$

- 原理的には、全てのCFG規則をもつ PCFG
- Penn Treebank から抽出したそのままの PCFG より高精度を達成する



# 非終端記号の細分化(1/2)

- 語彙化: Head percolation table [Magerman 1995] を用いて、非終端記号に head word を付与



Head percolation table

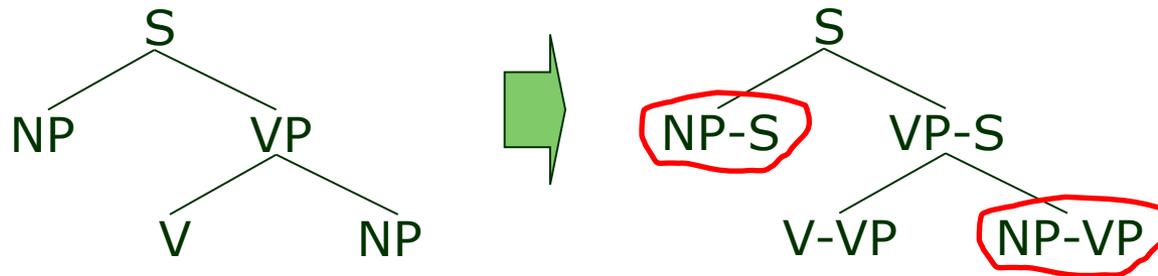
親の記号	主辞になる子の記号
S	VP, ...
VP	VP, VBD, VBZ, ...
NP	NN, ...
PP	IN, ...

Charniak [1996]: 80% vs. Magerman [1995]: 86%

- (参考) 語彙化の意味 [Gildea 2001; Bikel 2004]

# 非終端記号の細分化(2/2)

- 非終端記号だけでは構造を決める情報が少ない
- (例) 親の非終端記号で細分化 [Johnson 1998]



- 主語のNPと目的語のNPが区別できる
  - 主語は代名詞が出やすい
  - 目的語は長くなりやすい
- その他、様々な周辺情報で細分化 [Charniak 2000; Klein et al. 2003]

# より深い構造の抽出

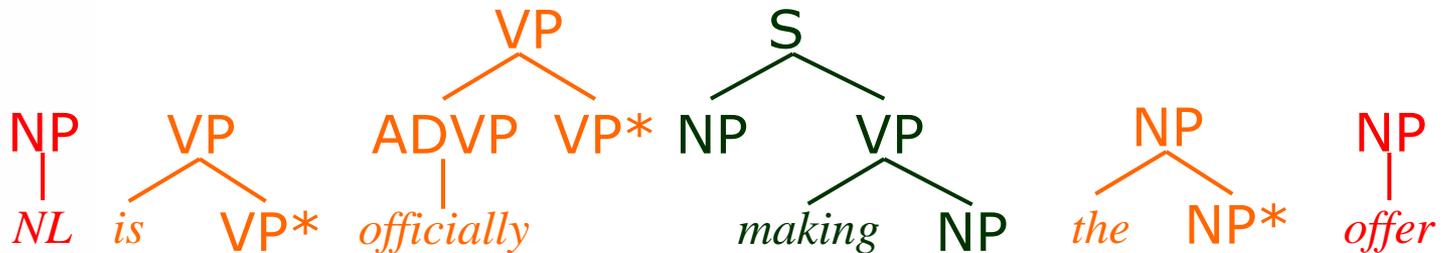
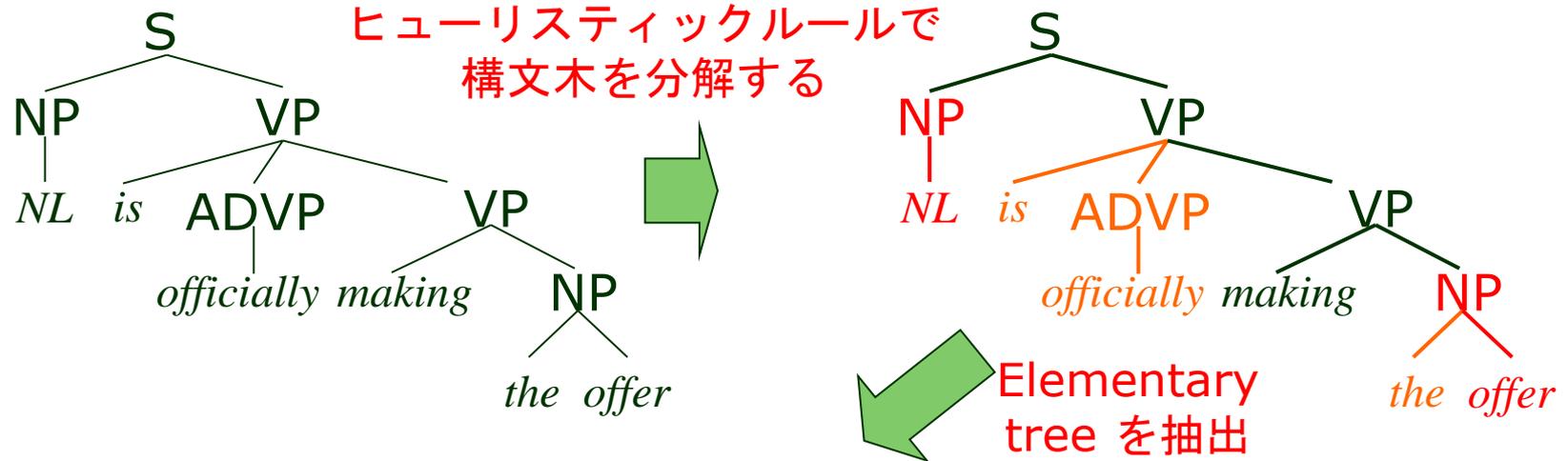
- CFGより深い構文構造や意味構造がほしい
- より深い構造のツリーバンクを作る？
- 非現実的：
  - 高コスト
  - 構造が複雑になると、矛盾・間違いが多発

→ Penn Treebank から、より高度な文法を自動抽出できないか？



# LTAG文法の自動抽出

- 構文木から LTAG の elementary tree を抽出 [Xia 1999; Chen et al. 2000; Chiang 2000]

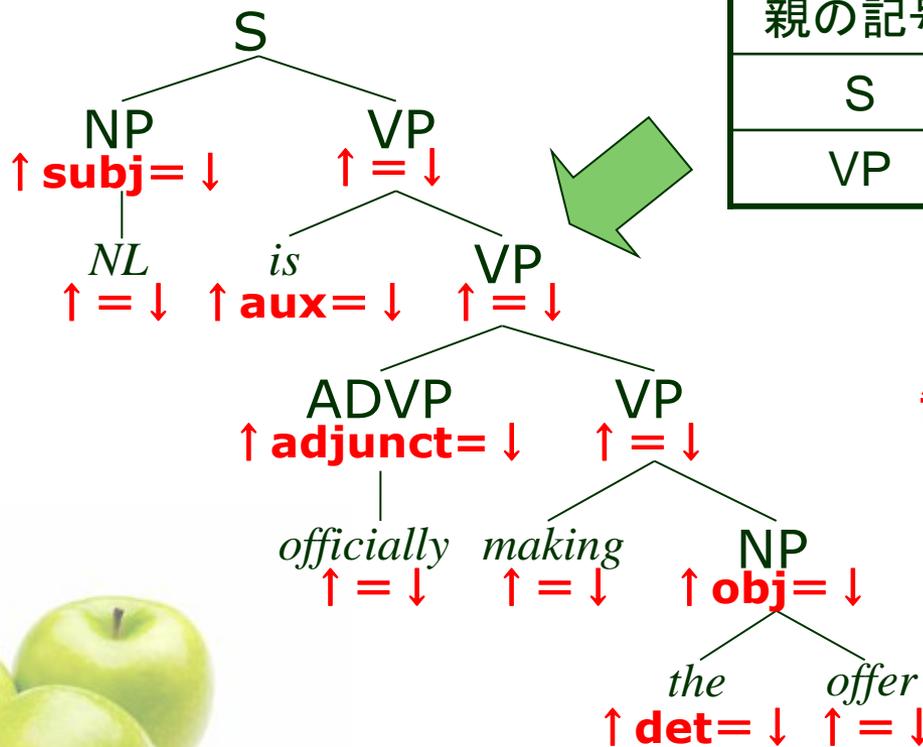


# LFG文法の自動抽出

- 構文木に f-structure を自動付与する [Cahill et al. 2002; Frank et al. 2003]

## 自動付与ルール

親の記号	子の記号
S	NP:↑subj=↓, VP:↑=↓
VP	NP:↑obj=↓, VP:↑=↓



## 制約解決

*f-structure*

```

[
  PRED make
  SUBJ NL
  OBJ [
    PRED offer
    DET the
  ]
  ADJUNCT officially
  AUX be
]
    
```

# 経験主義的文法の問題点(1/2)

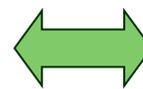
- ツリーバンク開発の問題
  - 正解の客観的基準が存在しない
  - 深い構造・複雑な構造の品質管理は困難

→ 文法理論に基づく合理的な品質管理・  
構造化が必要

こんなややこしい構造を  
書いてられない!



検証・開発



# 経験主義的文法の問題点(2/2)

## ● 自動的な文法抽出の妥当性

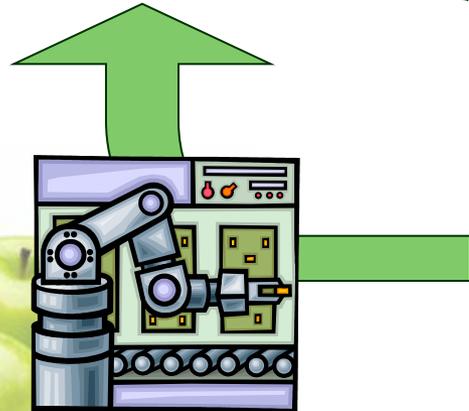


S → NP VP  
NP → DET N  
NP → N  
...

文法規則、辞書

自動抽出した LTAG, LFG 文法は正しいのか？

→ 文法を合理的に検証する必要



← コンピュータ



ツリーバンク



本当に正しい？



# まとめ (1/2)

- 合理主義的文法
  - 人手による文法規則と辞書の開発と中心とした文法開発
  - 合理的な利点
    - 言語学的な妥当性
    - 複雑な構造、深い構造の記述が容易
  - 問題点
    - 網羅性と一貫性のトレードオフ
    - 曖昧性解消の先送り
    - 性能評価の問題
- 経験主義的文法
  - ツリーバンクを中心とした文法開発
  - 経験的な利点
    - 網羅性
    - 一貫性
    - 機械学習・統計学習が容易
    - 評価も容易
  - 問題点
    - 正解の客観的基準が存在しない
    - 深い構造・複雑な構造の品質管理は困難
    - 自動的な文法抽出の妥当性

文法とツリーバンクの両方を開発することが重要!

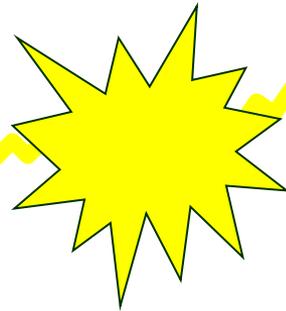
# まとめ (2/2)

- 経験主義的文法開発と合理主義的文法開発



経験主義的文法開発

直感＋アノテーション  
ガイドライン



合理主義的文法開発

辞書と文法規則による  
文法理論

# 参考文献

- H. Alshawi (Ed.) (1992) The Core Language Engine. MIT Press.
- A. K. Joshi and Y. Schabes (1997) Tree Adjoining Grammars. in G. Rosenberg and A. Salomaa, (eds.), Handbook of Formal Languages, vol. 3, pp. 69-124.
- XTAG Research Group (2001) A lexicalized tree adjoining grammar for English. Technical Report IRCS-01-03, University of Pennsylvania.
- A. Abeillé and M.-H. Candito and A. Kinyon (2000) FTAG: developping and maintaining a wide-coverage grammar for French. ESSLLI-2000.
- J. Bresnan (1982) The Mental Representation of Grammatical Relations. MIT Press.



# 参考文献

- S. Riezler, T. H. King, R. S. Crouch, J. T. Maxwell, R. M. Kaplan (2002) Parsing the Wall Street Journal using a lexical-functional grammar and discriminative estimation techniques. In Proc. of ACL 2002.
- R. M. Kaplan, S. Riezler, T. H. King, J. T. Maxwell, A. Vasserman (2004) Speed and accuracy in shallow and deep stochastic parsing. In Proc. of HLT/NAACL-2004.
- M. Forst and C. Rohrer (2006). Improving coverage and parsing quality of a large-scale LFG for German. In Proc. of LREC 2006.
- C. Pollard and I. A. Sag (1994) Head-Driven Phrase Structure Grammar. University of Chicago Press.



# 参考文献

- S. Müller (1996) The Babel-System – An HPSG Prolog Implementation. In Proc. of 4<sup>th</sup> International Conference on the Practical Application of Prolog, pp. 263 – 277.
- M. Siegel and E. M. Bender (2002) Efficient Deep Processing of Japanese. In Proc. of the 3rd Workshop on Asian Language Resources and International Standardization. COLING 2002 Post-Conference Workshop.
- G. Bouma, G. van Noord, R. Malouf (2000) Alpino: Wide-coverage Computational Analysis of Dutch. Computational Linguistics in the Netherlands. Selected Papers from the 11th CLIN Meeting.
- J. Carroll and T. Briscoe (2002) High Precision Extraction of Grammatical Relations. In Proc. of COLING 2002.



# 参考文献

- M. Butt, H. Dyvik, T. H. King, H. Masuichi, and C. Rohrer (2002) The Parallel Grammar Project. In Proceedings of COLING-2002 Workshop on Grammar Engineering and Evaluation. pp. 1-7.
- D. Flickinger (2002) On building a more efficient grammar by exploiting types. In Stephan Oepen, Dan Flickinger, Jun'ichi Tsujii and Hans Uszkoreit (eds.) Collaborative Language Engineering. Stanford: CSLI Publications, pp. 1-17.
- E. M. Bender, D. Flickinger, and S. Oepen (2002) The grammar Matrix. An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammar. In Proc. of the Workshop on Grammar Engineering and Evaluation at COLING 2002.



# 参考文献

- T. Götz and D. Meurers (1997) The ConTroll System as Large Grammar Development Platform. ``ENVGRAM" ACL-Workshop.
- A. Copestake and D. Flickinger (2000) An open-source grammar development environment and broadcoverage English grammar using HPSG. In Proc. LREC-2000.
- S. Oepen and J. Carroll (2000) Performance profiling for parser engineering. Natural Language Engineering, 6 (1) (Special Issue on Efficient Processing with HPSG):81-97.
- T. Baldwin, E. M. Bender, D. Flickinger, A. Kim, and S. Oepen (2004) Road-testing the English Resource Grammar over the British National Corpus. In Proc. LREC 2004, pages 2047-2050.



# 参考文献

- A. Frank, T. H. King, J. Kuhn, J. Maxwell (1998) Optimality Theory Style Constraint Ranking in Large-scale LFG Grammars. In Proc. of the 3<sup>rd</sup> LFG Conference.
- M. Marcus, B. Santorini, Marcinkiewicz (1993) Building a large annotated corpus of English: the Penn Treebank. Computational Linguistics 19.
- A. Bies, M. Ferguson, K. Katz, R. MacIntyre, V. Tredinnick, Grace Kim, M. A. Marcinkiewicz, B. Schasberger (1995) Bracketing Guidelines for Treebank II Style Penn Treebank Project
- G. Sampson (1995) English for the computer. Oxford University Press.
- S. Brants, S. Dipper, S. Hansen, W. Lezius, and G. Smith (2002) The TIGER Treebank. In Proc. Workshop on Treebanks and Linguistic Theories.

# 参考文献

- J. Hajic (1998) Building a syntactically annotated corpus: The Prague Dependency Treebank. In Issues of Valency and Meaning.
- E. Hinrichs, J. Bartels, Y. Kawata, V. Kordoni, and H. Telljohann (2000) The Tübingen treebanks for spoken German, English, and Japanese. In W. Wahlster (ed.), *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer.
- EDR (1995) EDR 電子化辞書使用説明書第2版. Technical Report TR-045.
- 黒橋、長尾 (1997) 京都大学テキストコーパス・プロジェクト. 言語処理学会第3回年次大会発表論文集.
- 前川、籠宮、小磯、小椋、菊池 (2000) 日本語話し言葉コーパスの設計. 音声研究 4-2.

# 参考文献

- E. Charniak (1996) Tree-bank Grammars. Technical Report CS-96-02, Department of Computer Science, Brown University.
- E. Charniak (1997) Statistical parsing with a context-free grammar and word statistics. In Proc. 14<sup>th</sup> National Conference on Artificial Intelligence.
- S. Sekine and R. Grishman (1995) A Corpus-based Probabilistic Grammar with Only Two Non-terminals. In IWPT '95.
- B. Carpenter and C. Manning (1997) Probabilistic parsing using left corner language models. In 5<sup>th</sup> IWPT.
- D. Magerman (1995) Statistical decision-tree models for parsing. In Proc. 33<sup>rd</sup> ACL.

# 参考文献

- A. Krotov, M. Hepple, R. Gaizauskas, Y. Wilks (1998) Compacting the Penn Treebank grammar. In Proc. 17<sup>th</sup> COLING.
- A. Krotov, M. Hepple, R. Gaizauskas, Y. Wilks (1999) Evaluating two methods for Treebank grammar compaction. Natural Language Engineering 5(4).
- M. Collins (1996) A new statistical parser based on bigram lexical dependencies. In Proc. 34<sup>th</sup> ACL.
- M. Collins (1997) Three generative lexicalised models for statistical parsing. In Proc. 35<sup>th</sup> ACL.



# 参考文献

- E. Charniak (2000) A maximum-entropy-inspired parser. In Proc. NAACL-2000.
- M. Johnson (1998) PCFG models of linguistic tree representations. Computational Linguistics 24(4).
- D. Gildea (2001) Corpus variation and parser performance. In Proc. 2001.
- D. Bikel (2004) Intricacies of Collins' parsing model. Computational Linguistics 30(4).
- D. Klein and C. Manning (2003) Accurate unlexicalized parsing. In Proc. ACL 2003.



# 参考文献

- F. Xia (1999) Extracting tree adjoining grammars from bracketed corpora. In Proc. 5<sup>th</sup> NLPRS.
- J. Chen and K. Vijay-Shanker (2000) Automated extraction of LTAGs from the Penn Treebank. In Proc. 6<sup>th</sup> IWPT.
- D. Chiang (2000) Statistical parsing with an automatically-extracted tree adjoining grammar. In Proc. 38<sup>th</sup> ACL.
- A. Cahill, M. McCarthy, J. van Genabith, and A. Way (2002) Parsing with PCFGs and automatic f-structure annotation. In Proc. 7<sup>th</sup> International Lexical-Functional Grammar Conference.



# 参考文献

- A. Frank, L. Sadler, J. van Genabith, and A. Way (2003) From treebank resources to LFG f-structures: Automatic f-structure annotation of treebank trees and CFGs extracted from treebanks. In A. Abeille (ed), Building and Using Syntactically Annotated Corpora. Kluwer Academic Publishers.
- T. H. King, R. Crouch, S. Riezler, M. Dalrymple, and R. Kaplan (2003) The PARC 700 Dependency Bank. In Proc. LINC 2003.
- S. Oepen, K. Toutanova, S. Shieber, C. Manning, D. Flickinger, and T. Brants (2002) The LinGO Redwoods Treebank: Motivation and preliminary applications. In Proc. COLING 2002.



# 参考文献

- F. Bond, S. Fujita, C. Hashimoto, K. Kasahara, S. Nariyama, E. Nichols, A. Ohtani, T. Tanaka, S. Amano (2004) The Hinoki Treebank: A treebank for text understanding. In IJCNLP-04.
- K. Toutanova, C. Manning, and S. Oepen (2002) Parse ranking for a rich HPSG grammar. In Proc. TLT2002.
- J. Hockenmaier and M. Steedman (2002) Acquiring compact lexicalized grammars from a cleaner treebank. In Proc. 3<sup>rd</sup> LREC.
- Y. Miyao, T. Ninomiya, and J. Tsujii (2004) Corpus-oriented grammar development for acquiring a Head-Driven Phrase Structure Grammar from the Penn Treebank. In Proc. IJCNLP-04.