

# 知的情報処理(10) 言語とは何か

慶應義塾大学理工学部  
櫻井彰人

## 言語の重要性

- コミュニケーションの道具として極めて重要
  - 古代より認識されていた
  - 旧約聖書創世記11章に現れるバベルの塔
  - 16世紀後半ネーデルランドにおいて非常に流行した主題、180点以上が現存。プリューゲルが有名



<http://www.h5.dion.ne.jp/~pieter/second-1.html>

## 旧約聖書創世記第11章（日本聖書教会「聖書 新共同訳」）

11:01 世界中は同じ言葉を使って、同じように話していた。11:02 東の方から移動してきた人々は、シナルの地に平野を見つけ、そこに住み着いた。11:03 彼らは、「れんがを作り、それをよく焼こう」と話し合った。石の代わりにれんがを、しっくい代わりにアスファルトを用いた。11:04 彼らは、「さあ、**天まで届く塔のある町を建て**、有名になろう。そして、全地に散らされることのないようにしよう」と言った。11:05 主は降って来て、人の子らが建てた、塔のあるこの町を見て、11:06 言われた。「**彼らは一つの民で、皆一つの言葉を話しているから、このようなことを始めたのだ**。これでは、彼らが何を企てても、妨げることはできない。11:07 我々は降って行って、**直ちに彼らの言葉を混乱させ、互いの言葉が聞き分けられぬようにしてしまう**。」

11:08 主は彼らをそこから全地に散らされたので、彼らはこの町の建設をやめた。11:09 こういうわけで、この町の名はバベルと呼ばれた。主がそこで全地の言葉を混乱(バラル)させ、また、主がそこから彼らを全地に散らされたからである。

[http://www.05.u-page.so-net.ne.jp/fd5/shinhiro/bible/gen11\\_36.htm](http://www.05.u-page.so-net.ne.jp/fd5/shinhiro/bible/gen11_36.htm)

## 言語とは何か

- “Language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of voluntarily produced symbols.” (E. Sapir : *Language*, 1921)
- “Language is the institution whereby humans communicate and interact with each other by means of habitually used oral-auditory arbitrary symbols.” (R.A. Hall : *An Essay on Language*, 1968)
- 「言語とは、人間の集団の中で、対人間のコミュニケーションにおいて使用され、または使用され得る、人間の環境の中での事象や過程を分析しつくす、任意の音声の、または音声の系列からなる構造化された体系である」(J.B. Carroll : *The Study of Language*, 1952, 『言語学と関連領域』大修館書店)

<http://lapin.ic.h.kyoto-u.ac.jp/soron/handout1.pdf>

注:手話が忘れられている

## 言語とは何か

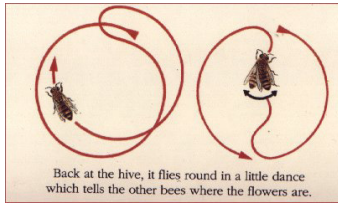
- コミュニケーション道具は言語以外にも多数
  - いるか・鯨の鳴き声、人間の表情、蜂のダンス、、、
- 音声とは限らない: 手話
  - しかしジェスチャは言語ではない
- 人間しか獲得・利用していない
  - チンパンジーやボノボに言語を学習したという主張はある。しかし、鼻真目にみても、単語を覚えたのみ
  - 人間はどんな(所謂未開の)種族でも言語を有する
  - 乳幼児は、非常に劣悪な環境でも、言語がある環境で育てれば、言語を非常な速度で獲得する

## 言語の特徴

- Hockett による13の特徴から
  - Hockett, Charles. 1960. *The Origin of Speech*.
- 意味性: 記号とその意味とが対応
- 任意性: その対応はまったく随機的
- 遠隔性: その時その場にはないものの記述が可能
- 生産性:
  - 新規な表現を聞き手が分かるように表現可能
- 伝承性: 親から子へ
- 二重性: 無意味記号から有意味記号の構成

## 蜂のダンス

- 蜂の動作の軸
  - 太陽のある方向に対する餌のある方向
- ダンスの(繰り返しまでの)長さ
  - 巣から餌までの距離
- 尻尾の揺れの速度
  - 餌の質
- これは言語か？



<http://www.hunnybee.com.au/hunny44.html>

## 自然言語と人工言語

- 次のものは言語か？自然言語か？
  - 日本語、英語、タガログ語、スワヒリ語、、、
  - エスペラント、アフリカンス、
  - JSL, ASL,,,
  - Algol, Fortran, C, Pascal,,,

## 手話は自然言語か？

日本手話と日本語対応手話とは別

- ASL, BSL等,手話(Deaf Sign Language)として121言語が数えられている

[http://www.ethnologue.com/show\\_family.asp?subid=90008](http://www.ethnologue.com/show_family.asp?subid=90008)

言語の要件は満たす。その他

独自の文法、伝達速度は音声言語と同等

人工言語ではない(文法は自律的発達)

脳の活動部位は同じ(日本語/日本手話. 母語の場合)

手話にも、音声言語と同様の失語症がある

Sakai, K. L., Tatsuno, Y., Suzuki, K., Kimura, H. & Ichida, Y.: Sign and speech: Amodal commonality in left hemisphere dominance for comprehension of sentences. *Brain* 128, 1407-1417 (2005).

## 言語の数

- L.H. Gray, *Foundations of Language*, 1939 死語を含め、方言は除いて2,796の言語があるとされている。
- B.F. Grimes, *Ethnologue 1978* 死語を除いて5,103の言語があるとされている。
- 『言語学大辞典』(三省堂) 3,500の言語について記述があるが、名前だけで記述のない言語がそれ以外にさらに3,500ある。
- Voegelin & Voegelin, *Classification and Index of the World's Languages*, 1977, New York, Elsevier には20,000にのぼる言語と方言が記載されている。 <http://lapin.ic.h.kyoto-u.ac.jp/soron/handout1.pdf>
- [http://www.ethnologue.com/language\\_index.asp](http://www.ethnologue.com/language_index.asp)には、7,413の主たる言語名がリストされている (2002年には6800であった。これは言語が増加していることを意味しない。現実には、どんどんなくなっていつている)

どこまで違えば別の言語か？ 沖縄語は？大阪弁は？津軽弁は？ケセン語は？

## 使用人口による言語ベスト20

1. 中国語	1,000	11. フランス語	70
2. 英語	350	12. パンジャブ語	70
3. スペイン語	250	13. ジャワ語	70
4. ヒンディー語	200	14. ビハーリー語	65
5. アラビア語	150	15. イタリア語	60
6. ベンガル語	150	16. 朝鮮語	60
7. ロシア語	150	17. テルグ語	55
8. ポルトガル語	135	18. タミル語	55
9. 日本語	120	19. マラーティー語	50
10. ドイツ語	100	20. ベトナム語	50

(単位は百万人)

<http://lapin.ic.h.kyoto-u.ac.jp/soron/handout1.pdf>

## 日本語は特殊な言語か？

- 外国人にとって日本語は他のことより難しい
  - そう思う79% そう思わない11% どちらとも言えない7%
- 日本語のような情感の豊かなことばは外国語には翻訳できない。
  - そう思う71% そう思わない14% どちらとも言えない11%
- 英語は論理的だが、日本語は論理的でない。
  - そう思う31% そう思わない40% どちらとも言えない19%

<http://lapin.ic.h.kyoto-u.ac.jp/soron/handout3.pdf>

注：言語とその文字表記、言語とその運用について考えよ

## 日本語は複雑・特殊な言語か？

- 世界で最も多い母音体系は5~7母音
  - ドイツ語15母音, フランス語16母音, 英語21母音
- 世界の言語で最も多い子音
  - p b t d C k g f s tC m n N ð w l r j h
- SOV 39%, SVO36%, VSO15%, VOS5%, SV/OVS 5%
- 疑問小辞の位置 (Cl. アジェージュ『言語構造と普遍性』白水社)
  - 文頭は稀で文末が多い。(フランス語は例外的*Est-ce que tu es prêt?*)
  - 文頭に来るのはSVO言語に多い。
  - 文末に来るのはSOV・SVO両方に見られる。
  - 後置詞言語の61.1%は文末に来る。

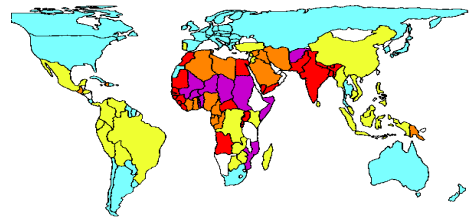
角田太作『世界の言語と日本語』くろしお出版  
<http://lapin.ic.h.kyoto-u.ac.jp/soron/handout3.pdf>

## 日本語は不完全な言語か？

- 「この国語はある点で不完全なものである。何となれば、名詞は格による変化を欠き、単数複数別の別および性の区別をもたず、動詞は人称および単数複数の別を欠き、その他にもヨーロッパの言語にはみられない欠陥があるからである。」(ロドリゲス『日本大文典』1604~1608)
- 「吾々は子供から今の国語に慣らされ、それ程に感じていないが、日本の国語ほど不完全で不便なものはないと思ふ。その結果、いかに文化の進展が阻害されていたかを考へると、これは是非ともこの機会に解決しなければならぬ大きな問題である。(…)私は60年前、森有礼が英語を国語にしようとした事を此戦争中、時々想起した。若しそれが実現していたら、どうであつたらうと考えた。日本の文化が今よりも遙かに進んでいてであらう事は想像できる。(…)そこで私はこの際、日本は思ひ切つて世界中で一番いい言語、一番美しい言語をとって、そのまま、国語に採用してはどうかと考へている。それにはフランス語が最もいいのではないかと思ふ。(…)フランス語を想つたのは、フランスは文化の進んだ国であり、小説を読んで見ても何か日本人と通ずるものがあると思はれるし、文人達によって成る時、整理された言葉だともいふし、そういう意味でフランス語が一番よきさうな気がするのである。」(志賀直哉『改造』昭和21年4月号)

## 言語と文字

- 誰でも言語は使える。しかし文字が使えるとは限らない
- ひらがな・カタカナ・漢字と表記が複雑な日本でも、非識字率は1%以下。



非識字率  
 95%以上 90%以上 80-90% 70%以上 60%以上 50%以上 40%以上 30%以上 20%以上 10%以上 10%以下 不明  
<http://www.nic-nagoya.or.jp/hskzkzritu.htm>

## 日本語は不完全な言語か？

- 主語がなく、従って、非論理的である
  - 今日はカゼなので出席できません
  - 初めてお手紙差し上げます
- 主語が二つ！
  - 象は鼻が長い、僕は君が好きだ
- 私は人間である！
  - 僕はウナギだ。僕はキツネ。こっちはタヌキ
  - 僕は並。

## 言語学からのアプローチ

- 言語の三要素
  - 統語論
  - 意味論
  - 音韻論
- いずれも、人間と他の動物との差異を際立たせている
- 以下では、特に統語論(構文論)について考える

## 江戸時代の識字率

幕末期の日本全国の識字率(読み書きできる人の全人口に占める割合)についてはいろいろな研究があるが、武士階級はほぼ100%、庶民では男子54%、女子19%というのから、寺小屋の生徒数から推定した男子49%、女子21%までさまざまである。江戸府内だけに限定すれば、男女とも70%から80%には達していたし、更に、府内の農村部を除外した市街地に限定すれば、90%以上であった。これに対して、同時期のイギリスでは、人口の大きな部分を占める下層庶民の場合、ロンドンでさえ辛うじて字の読める子どもは10%に達せず、書ける子どもとなればその数分の一だった。

石川英輔著『雑学「大江戸庶民事情」』講談社文庫より)

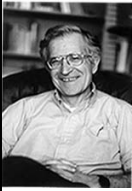
## 構文論の誕生

- 音韻論と意味論があるのは当然
  - ことばには(少なくとも)音と意味がある
- 形態論や語彙論があっても当然
  - 言語音→形態素→単語→単語の類型
- 文法は「学校で生徒に「正しい」ことばの使い方を教えるための方便」であった(20世紀前半まで)
  - 規範文法(prescriptive grammar)と呼ばれる
  - 規範文法は、普通にことばを話す人の実感とかけ離れていた
  - 規範を作ることから客観的に観察できることをできるだけ正確に記述することへ移る
  - しかし、Chomskyの生成文法理論の提案により大きく変わる

## 規範文法の例

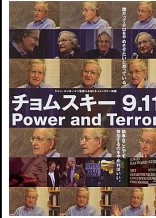
- Don't end a sentence with preposition
  - What is she talking about
  - vs. About what is she talking
- Don't use double negation
  - I didn't see nothing/I ain't seen nothing
  - vs. I didn't see anything
- Use whom instead of who when this word is an object rather than a subject
  - Who did you meet
  - vs. Whom did you meet

## Avram Noam Chomsky



- 1928 Philadelphia生まれ
- 1955 Pennsylvania大学から言語学で学位
- 1961 MIT教授、大学院・言語学科設立
- 1988 京都賞(基礎科学分野)受賞
  - 「生成文法理論」を提唱して、言語学の革命的な転回を引き起こし、これを通じて人間の精神構造を解明するという野心的なプログラムを可能にした。
  - *Syntactic Structures* (1957), *Aspects of the Theory of Syntax* (1965), *Lectures on Government and Binding* (1981), and *The Minimalist Program* (1995).

## Avram Noam Chomsky



- ベトナム戦争以来米国の、外交を中心とした政治・政策、メディアの痛烈な批判者
- 飽くなき反抗者(“rebel without a pause”)
  - by Bono Vox, full-time lead singer of U2
- 主な著作
  - *American Power and the New Mandarins* (1967), *Peace in the Middle East* (1974), *The Fateful Triangle* (1983) *Deterring Democracy* (1991),

## しかし、これは想定外だろう



20日、国連総会一般討論で、チョムスキー氏の著書を手元に演説するベネズエラのチャベス大統領(ライター)

<http://www.sankei.co.jp/news/060921/kok002.htm>

## 生成文法研究

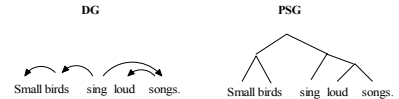
- 言語研究は言語を生み出すもとなる人間の精神(脳)内の言語知識の解明を目的としてこそ、価値がある
- 課題
  - 人間の持っている言語知識とはどのようなものなのか、
  - その知識はどのようにして獲得されるのか、
  - 獲得された知識はどのように使用されるのか、
  - 言語知識の獲得を可能にする生物学的基盤はどのようなものなのか

## 生成文法

- 特定の言語の分析・記述ではなくて話し手が従来の言語能力によって**文法的な文のみを限りなく生成していく仕組み**を文法と考え、深層に仮定される抽象的な基本構造から、変形によって現実の文の構造に至る規則の体系(後半部分は後に修正される)(広辞苑)

## 句構造文法

- 文を句単位に分けて、句の(埋込み)構造で構文を表現する
  - 語と語がどのように関係しているかで構文を表現する方法もある。依存文法(dependency grammar)という
- 表現方法
  - 非終端記号(書き換えられる記号)、終端記号、開始記号、および書換え規則
- 極めて一般的な枠組み。情報科学に大きな影響



## 自然言語処理とは

- 立場によって異なる
- 工学的には: 自然言語を処理するシステムを作る
  - 情報検索システムを作る
  - コールセンターを作る
  - Q&Aシステムを作る
- 心理学的には: 人間の自然言語処理能力を理解する
  - 言語をどう獲得しているのか
  - 言語をどう理解しているのか
  - 言語をどう生成しているのか
  - 言語はどう思考と結びついているのか

## チューリングテスト

- Turing, A (1950). Computing machinery and intelligence. Mind, 59, 433-460.
- 問題
  - I propose to consider the question, 'Can machines think?'
- ものまねゲーム
  - The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'.
- チューリングテスト
  - We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?'



Alan Turing

<http://www.ioebner.net/Prize/TuringArticle.html>

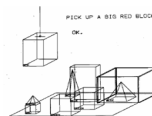
## 自然言語処理史

- 1950年代: 機械翻訳 夢の時代
  - 機械翻訳プロジェクトが多数
- 1960年代: 挫折の時代
  - ALPAC report (1966)
    - 主たる目的であるロシア語翻訳の需要は供給以上ではない
    - 事後編集するも、直接翻訳以上にコストがかかる
- 1970年代: 人工知能研究の進展
  - 動く対話システム (例: SHRDLU)
  - 記号処理システム (例: LISP)
  - 情報検索システム

<http://ourworld.compuserve.com/homepages/wjhutchins/Alpac.pdf>  
<http://www.nap.edu/books/ARC000005/html/index.html>

## 対話システム: SHRDLU

Terry Winograd. 1971. MIT Ph.D. Thesis.



- 辛口批評:
  - This is one of the only dialogs it knows.
  - SHRDLU is too stupid to make mistakes.
  - Beautiful "Demo-ware"

Person: PICK UP A BIG RED BLOCK.  
Computer: OK. (does it)  
Person: GRASP THE PYRAMID.  
Computer: I DONT UNDERSTAND WHICH PYRAMID YOU MEAN.  
Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.  
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.  
Person: OK. (does it)  
Person: WHAT DOES THE BOX CONTAIN?  
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.  
Person: WHAT IS THE PYRAMID SUPPORTED BY?  
Computer: THE BOX.  
Person: HOW MANY BLOCKS ARE NOT IN THE BOX?  
Computer: FOUR OF THEM.  
Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?  
Computer: YES. THE RED CUBE.



Terry Winograd

<http://hci.stanford.edu/~winograd/shrdlu/>

## 自然言語処理史(2)

- 1980年代: 合理主義
  - 構文論、意味論に注目
  - 文法 (LFG, GPSG) や知識 (KL-ONE, CYC, 等) における論理的枠組み
  - すべてが、手作り
  - スケールアップしないし、ロバストでない



Joan Bresnan (LFG)



Gerald Gazdar (GPSG)

## 知識ベース: CYC

- CYC 流 "every animal has a mother"  
(forAll ?A  
(implies  
(isa ?A Animal)  
(thereExists ?M  
(and  
(mother ?A ?M)  
(isa ?M FemaleAnimal))))))
- 世界にある知識を矛盾なく記述することはできない
- 知識の保守は、悪夢・悪夢・悪夢
- しかし、今でも、存在し、発展している。その要因は、“Semantic Web” であり、自然言語処理でいえば WordNet である
- 参照: <http://opencyc.org>



Doug Lenat

## 自然言語処理史(3)

- 1990年代~2000年代: 経験主義
  - より明確な課題、例えば品詞付けや浅い構文解析、に注力 (Penn TreeBank)
  - その範囲内でならロバストなシステム
  - 実データに基づく評価
  - 記号主義、しかし、統計的アプローチ
- 未来は？
  - web を梃子に新たなアプリがざくざく
  - 技術のコベルニクスの転回はあるか？

## 機械翻訳を試してみよう

- 例: リンク集
  - <http://www.dio.ne.jp/user/bestsites/translate.html>
  - 大阪語への変換  
<http://www.yansite.net/osakaproxy.html>

## 時間があつたときの即レポ

- 機械翻訳ソフトを使い、機械翻訳の得手・不得手について論じてください。
  - 簡単な日本語文、複雑な日本語文を例文とする。
  - 翻訳結果を日本語に翻訳しなおしてみてください。どうなりますか？
- 例文(単純な構文)
  - 我輩は猫である。名前はまだない。
  - 英語はたいへん難しい言葉です。何しろ勉強を始めたのが中学校ですが、まだ、喋れません。もともと文法は得意です。
  - 困ったよ。もう期末なのに、まだ予算未達なんだ。
  - 僕はうなぎにするよ。君はどうする？
- 例文(複雑な構文)
  - 太郎が買ったケーキを次郎が食べてしまった。
  - これはジャックの立てた家に置いた小麦を食べたねずみを殺した猫を苛めた犬です。
- その他皆さんが考えた文で試してみてください

自動翻訳サイトのリンク集: <http://www.dio.ne.jp/user/bestsites/translate.html>

## 構文法

- 当該言語の文法的な文をすべて生成し、非文法的な文を一切生成しないような装置  
(Chomsky 1957)
  - 装置: 人間の話し手と同じような方法で文を生成するモデル (Lees 1957)
- 未完成、しかし多くの成果
- その基本部分が句構造文法

## 句構造文法と書き換え規則

- 文法とは、
  - 語彙(終端記号)
  - 文法範疇(非終端記号、品詞)
  - 書き換え規則
  - 開始記号(文)
- の4要素からなる。
- 文の生成は、開始記号を最初の対象として、書き換え規則を次々に適用し、全ての文法範疇を語彙に置き換える過程である
- 文の構文解析は、文の要素(形態素)を語彙にマッチさせ、語彙から書き換え規則を逆に適用することによって文法範疇に書き換え、最後に開始記号にたどり着く過程である。

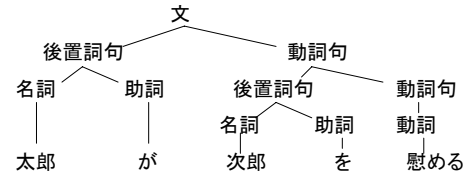
## 文法の例

- 語彙:
  - 太郎、花子、次郎、が、を、と、慰める
- 文法範疇:
  - 名詞、動詞、助詞、名詞句、後置詞句、動詞句、文
- 開始記号: 文
- 書き換え規則
  - 名詞 → 太郎 | 花子 | 次郎
  - 助詞 → が | を | と | は
  - 動詞 → 慰める
  - 後置詞句 → 名詞 助詞 | 名詞句 助詞
  - 名詞句 → 名詞
  - 動詞句 → 動詞 | 後置詞句 動詞句
  - 文 → 後置詞句 動詞句

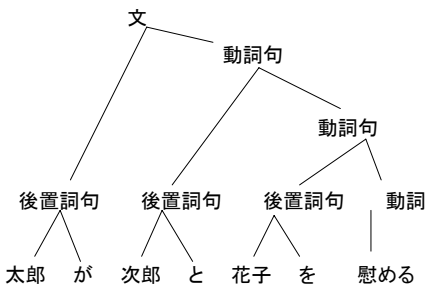
1. 文→後置詞句 動詞句
2. 文→名詞 助詞 動詞句
3. 文→太郎 が 動詞句
4. 文→太郎 が 後置詞句 動詞句
5. 文→太郎 が 名詞 助詞 動詞句
6. 文→太郎 が 次郎 を 動詞
7. 文→太郎 が 次郎 を 慰める

## 構文構造の表現

- 括弧
  - (太郎が((次郎を)慰める)))
- 構文木



### ■太郎が次郎と花子を慰める

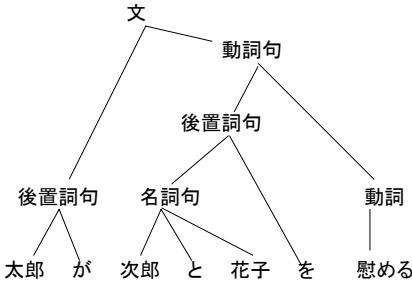


## 簡単ではない例

- 太郎が買ったケーキを食べた
- 太郎が次郎と花子を慰める
- I saw a girl with a telescope.
- 同僚が東京へ転動するので送別会をしてくれた
- Since Jay always walks a mile seems like a short distance to him.  
→ 袋小路文(garden path sentence)

■ 太郎が次郎と花子を慰める

名詞句 → 名詞 と 名詞



## 形式言語

■ 形式言語と自然言語

- 形式言語: Java や Prolog のように人間が発明し厳密に定義した言語
- 自然言語: 日本語、中国語、英語、日本語等
- 形式言語は自然言語の原始的な近似と考えることができる。特に、構文法。

## 形式言語のBNF記法による定義

■ BNF記法 (Backus Naur Form) の4つの部品

- 終端記号 (terminal symbol): 対象言語の文字列を構成する単位文字 (実際は、単語のことも)
- 非終端記号 (non-terminal symbol): 対象言語の節や句に相当する部分を表現する記号
- 開始記号 (start symbol): 対象言語の完全な文に対応する記号
- 書換え規則 (rewrite rule): 左辺 → 右辺; 左辺は1個の非終端記号、右辺は終端・非終端記号の0個以上の列

■ さきほどの定義と全く同じ

## BNFによる形式言語定義例

- $Exp = Exp \ Op \ Exp \mid ( \ Exp \ ) \mid \text{Num}$   
 $Num = \text{Digit} \mid \text{Num} \ \text{Digit}$   
 $\text{Digit} = 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$   
 $Op = + \mid - \mid * \mid /$
- 開始記号が  $Exp$  の時、この文法はどんな言語を定義しているのであろうか？

## より一般的な定義方法(1)

■ 前記4材料と、書換え規則の変種

- 0型文法 (帰納的可算文法): 書換え規則の両側に任意個数の終端・非終端記号
- 1型文法 (文脈依存文法): 右辺の記号個数は左辺の個数以上
- 2型文法 (文脈自由文法, CFG): 左辺は単一の非終端記号
- 3型文法 (正規文法): 左辺は単一の非終端記号、右辺は単一の終端記号と0個または1個の非終端記号

## より一般的な定義方法(2)

クラス	書換え規則の例	言語の例
帰納的可算	$A \ B \ \rightarrow \ C$	(原理的には) any
文脈依存	$a \ X \ B \ \rightarrow \ a \ b \ Y \ B$	$a^n b^n c^n$
文脈自由	$S \ \rightarrow \ a \ S \ b$	$a^n b^n$
正規	$S \ \rightarrow \ a \ S$	$a^* b^*$

## より一般的な定義方法(3)

- BNF記法はCFGに相当
- 自然言語は、CFGより複雑
  - Javaでも「型宣言なしの変数使用は禁止」はCFGでは書けない
- しかし、文脈依存文法や帰納的可算文法では表現しにくい
- けれども、骨格はCFGで記述できる
- 様々な方法(CFG+ $\alpha$ )が考案されている

## Chomskyの目指すもの

- 「言語能力」という脳に内在している認知能力の理論的説明
  - 言語行動から言語知識という心的表示へ
  - 普遍文法が存在
- 記述的妥当性と説明的妥当性

## 普遍文法(仮説)

- “the system of principles, conditions, and rules that are elements or properties of all human languages... the essence of human language” [Chomsky, 1975]
- 遺伝的に、人間のみ、付与されている
- 個別言語の構文法の母型
  - 個別文法は、パラメータの差異のみ
  - 例: 文の変形は句を単位として行われる
  - 例: 名詞と動詞とがある

## 言語獲得装置(仮説)

- 環境から与えられる言語例からその構文法を獲得する機構
  - 普遍文法のパラメータを設定する
  - 不完全な文、不十分な量の文から、急速に完全な構文法を獲得(プラトンの問題)
    - 計算論的には、強いバイアスなしには学習不可能
  - 幼児の一定時期のみ活動する(臨界期)
  - 遺伝的に規定され、人間のみにある

## 言語生得説対学習説

- 言語は生得的か後天的か。より正確には
- 言語の学習・獲得機構は、言語特有の獲得機構か、一般の学習機構か?
  - 生得説: Chomsky派
  - 学習説: Elman派
  - Science (1996, 1999)誌上の議論

## 手話文法発達の例

- 先天的聾者は、親(健聴者)のジェスチャーをもとに、文法(のもと)を構成
  - S.Goldin-Meadow and C.Mylander, spontaneous sign systems created by deaf children in two cultures, Nature 391, 279-281 (1998).
- Nicaragua 手話の発達
  - 内戦終了後、家庭にいた聾者(子供)が学校へ
  - 文法の統一、文法の複雑化
    - A.Senghas and M.Coppola, Children creating language: how Nicaraguan Sign Language acquired a spatial grammar, Psychological Science, 12, 323-328 (2001).
    - <http://www.unet.maine.edu/courses/NSLP/>
- 日本手話も同様の発達を経た(と考えられる)
  - 聾学校では禁止され、自立的発達せざるをえなかった

## 言語の機序: 究極の難問

- 脳はどのようにして言語を生み出すか？
- なぜ難問か？
  - 理由1: 言語は脳の高次機能の頂点にある
  - 理由2: 言語は人間のみにも備わった機能
  - 理由3: 言語は他のさまざまな認知機能と深く結びついている

## ピジン(pidgin)

- 2つ以上の言語の接触から生まれる言語(Contact Language)
- ネイティブ・スピーカーがいない
- 接触に関与したどの言語とも相互理解が不可能
- 文法は、もとの言語より単純であるのが普通
- 発展段階
  - 文法規則が定まらず語彙が極めて少ないジャーゴン(jargon)
  - 安定したピジン(stabilized pidgin)
  - 普通の言語といってもよいほど拡大したピジン(extended pidgin)
- 有名な例
  - Russenorsk: 18c~20c初頭。ノルウェー沖で操業するロシア人漁民とノルウェー人漁民
  - Chinese Pidgin English: 17c~19c半ばまで中国南部で
  - Fanakalo: 南アフリカ共和国の金鉱山付近で現在も。

<http://www.info.sophia.ac.jp/fs/fukusen/gengo/gen-0404.htm>

## クレオール(creole)

- ピジンがある集団の母語となったもの
  - このことを「クレオール化(creolization)」という
  - 母語: 集団の生活の必要をすべて満たす普通の言語
- ピジンの発展段階とクレオール
  - 初歩的なピジンから急激に
  - 安定化or拡大化したピジンが何世代も経て漸進的に
- 時制・法・アスペクトの表示法などいくつかの構造的な共通特徴
- 例

81のクレオールが上げられている

<http://www.info.sophia.ac.jp/fs/fukusen/gengo/gen-0404.htm>

## ピジンとクレオール

- ピジンとクレオールの言語的な違い
  - jargonとクレオールとは明確に違う、
  - 拡大ピジンとクレオールの比較は難しい
- いくつあるか？
  - だいたい100前後
- どんな言語で
  - 英語、フランス語、ポルトガル語、スペイン語、オランダ語の語彙
  - アフリカ諸語の語彙
  - アメリカ・インディアン諸語の語彙

<http://www.info.sophia.ac.jp/fs/fukusen/gengo/gen-0404.htm>

## 「始言語」を求める古代の試み

- Psammetichos (B.C. 7c): 古代エジプト王。赤ん坊を隔離して育てたところ、最初にプリュギア語で「パン」を意味する *bekos* という言葉を発した。
- James IV (A.D. 15c): <sup>ヘドトスによる</sup>スコットランド王。赤ん坊を隔離して育てたところ、最初に話したのはヘブライ語であった。

ロバート・リンゼイによるという

一般に、幼児期に親から離されることは非常なストレスである。しかも、言葉のない環境で育てられれば、満足に言語が獲得されることはない、とするのが現在の通説である。幼児虐待、聾者への偏見から現代でもときおり、幼児期の言語環境が劣悪な例が見られる。その結果からの知見である。

## 言語起源論のいろいろ

- Bow-wow theory ワンワン説: 動物の鳴き声などを真似た擬音語が言語の起源である
- Pooh-pooh theory プープー説: 痛みや怒りなどの感情・感覚が声となって外に出たものが言語の起源である
- Ding-dong theory リンリン説: 周囲の刺激に反応し、環境と調和する音声を発することから言語が発生した
- Yo-he-ho theory よいとまけ説: 共同で労働するときに、リズムをとったりかけ声をかけたりするのが言語の起源である (Otto Jespersenの分類)

デイヴィッド・クリスタル「言語学百科事典」

ゴシップ起源論 霊長類が社交的活動として行っているグルーミング(毛づくろい)に代わる手段として、言語は誕生し、発展していった。ロビン・ダンバー「ことばの起源」(青土社)