コーパスという生態系に語句の意味を探る

海洋生物学者が海に生物の生態を探るように

黒田 航

(独) 情報通信研究機構 けいはんな情報通信融合研究センター

1 はじめに

1.1 用法基盤主義の意義

言語現象の記述，説明に対する**用法基盤の姿勢** (usage-based approach) [1, 4] は言語の次の特徴を強く意識したものだと考えることができる:

(1) (複雑系としての) 言語には，どんなにすぐれた言語学者の直観も及ばない，微妙な性質が備わっている可能性がある．

これは言語の諸特徴を**文法** (grammar) という形で規則化，体系化することを至上の理念とする生成言語学の方法論 [2] への根本的な異議申し立てであり，生成文法によって歪められた言語観を是正し，言語学を経験科学的に変換する可能性をもった提案であった．

だが，一つのことを疑問にせざるを得ない: **提唱と受容から約 20 年後の今，用法基盤の理念はどれほどしっかり実践されているか?**

成立以来，認知言語学の手法は少なからず**直観基盤** (intuition-based) であった．この点で，認知言語学は，決戦を挑んだ相手である生成言語学と何も変らない．このため，両者の違いは，最終的には単なる説明用語やイメージの好みの違いに堕落する可能性がある．

認知言語学の枠組みであれ生成言語学の枠組みであれ，直観に過度に依存し作例を中心とした研究法は，次の理由から用法基盤アプローチの理念である (1) と**原理的に矛盾する**:

(2) **直観基盤の研究は確証バイアスを助長する**．

その理由は，作例中心の研究では，次の二点によってバイアスのかかった観察が助長されるからである:

(3) **データ収集へのバイアスの存在:** 確証バイアスによって収集されるデータが偏っている可能性があり，これはデータ収集が「行き当たりばったり」であることによって助長される．

(4) **直観へのバイアスの存在:** あらゆる言語研究者の直観は (好みや信念によって) バイアスされており．どんなにすぐれた言語学者の直観でも，それが正確に言語の実体を反映している保証はない[1)]．

言語を自然科学者の眼で観察し，記述し，説明するのが言語学者の仕事であるならば，確証バイアスは致命的な欠陥である．特に (4) と (1) との矛盾は致命的で，絶対に克服されなければならないものの一つである．

1.2 コーパス言語学の利点

確証バイアスからどう離脱するかという問題は，コーパス言語学では比較的自然な形で克服されている[2)]．

1.2.1 コーパス言語学を成立させる言語の基本的特徴

コーパス言語学が曲がりなりにも成立している以上，言語にはコーパス言語学を成立させる基本的な特徴がある，と考える必要がある．それが何であるかは，従来の研究では自明視されるか，不問になっている．これを明らかにするために，はじめに私は次の点を強調する:

(5) コーパス言語学を語彙分類学，その応用としての辞書編纂と同一視するのは，(無知に基づく) 単なる短絡であり，かつ偏見である．

(6) コーパス言語学を統計学と同一視するのは，(無知に基づく) 単なる短絡であり，かつ偏見である．

従来のコーパスを利用した辞書学，辞書編纂学，その基礎研究としての語彙分類の研究成果が示しているのは，語彙の使用の実態を調べるという目的のためにコーパスの利用が有効だということでしかない．それ以上の意味はない．**コーパスの利用可能性は，ほかにもまだまだある**．広義のコーパス言語学は，この種の情報に限られるわけではないし，限られるべきではない．

何百行，あるいは何千行もの KWIC 検索例を眺めるのがそれ自体で楽しみな研究者，緻密な語彙分類に満足し，それ以上の研究レベルに登らない研究者は，コーパス言語学の分野に，確かに多いような気がする[3)]．

---

1) 確証バイアスは，説明を志向する研究パラダイムに顕著である．

2) 実際，コーパス言語学がこれまでに明らかしてきた結果の幾つかは，しばしば認知言語学的な観点からすると反直観的なものである．例えば *give* の用法でもっとも頻度の高いのは軽動詞としての用法である (黒宮公彦による私信).

3) 世の中，分類が趣味の人々がいたって，ぜんぜんおかしくないのだ．それは健全な趣味で，他人にも有益な趣味である．実際，生物分類学にはそういう側面がある．

だが，仮にそれがコーパス言語学の現状の正しい観察だとしても，それはあくまでも現時点で，しかも典型的にそうだ，という以上のことではない．それをもってコーパス言語学の可能性を語るのは明らかにおかしい．典型例にはコーパス言語学のポテンシャルが完全に実現されているとは限らない．次の問いかけが重要である:

(7) **なぜコーパスはこの目的のために有用なのか?**

この根本的な問いに答えをもっている人は少ないように私には思われる．この問いの答えを部分的に §2 で提供したいと思うが，その前に下準備をしよう．

1.2.2 コーパス言語学は本当は何の研究であるべきか?

コーパス言語学の範囲を確定するのは難しいし，それをムリに限定することには，おそらく意味がない．極端な話をすれば，**コーパスを利用した言語学は，すべてコーパス言語学と呼ばれてしかるべき**である．

だが，コーパス言語学という研究手法の定着，研究分野の確立には，単に「コーパスを利用した言語研究」という以上の何か特別な意味が付与されている．端的には**データの定量的分析に基づく，語彙特性の客観的分析の手法という意味**が付与されている．

その結果には良い面と悪い面の両面がある．良い面としては，コーパス言語学は，従来の主観的解析方法 — 主に生成言語学の内観主義的方法 — の限界，あるいは無知を知らしめる効果が大であった．その反面は，コーパスの利用可能性に特定の利用法に限定されていた．

一番の問題はコーパス言語学が語彙特性の研究と同一視されている点である．この制限は取り外され，コーパス言語学の可能性は広げられなければならない，と私は痛感する．以下では，そのような拡張のための概念的基礎を敷くつもりである．コーパスを**語の生態系** (ecological system of words) の時間断面と見なすアナロジーの有効性を指摘し，それが用法基盤のアプローチの実践に最適なモデル化であると主張する．

1.3 複雑系としての言語, 生態系としての言語

だが，用法基盤主義は，認知言語学が単にコーパス言語学の猿真似をすれば実現される類のものではない．

コーパス言語学が得意とする用法の分類は，科学的な言語研究に不可欠な要素だが，それで十分というわけではない．用法の分類は科学的な言語研究の始まりであって，その終わりではない．科学的な言語学は，コーパス言語学が切り開いた地平を踏破しなければならない．

言語は単に**複雑系** (complex system) であるばかりではない．それは生態系，つまり “生きた系” でもある．この場合，言語というのは，異なる “種” の個体によって構成される体系であり，そこでの “種” とは “語” だと理解されることになる．

私は問題を一つ，重要な単純化している．ある生態系—特に進化する生態系— における種の定義は自明ではない．同様に，ある言語における語の定義は自明ではない．だが，生態系における種の概念と言語における語の概念が相同であるのが私のモデル化の基本である．これは仮説であるので，まちがっていれば，撤回する．

2 “語” の生態学としてのコーパス言語学

2.1 コーパスは語の生態系の時間断面である

私がコーパスを有意義に利用するにあたって根本的に重要だと思う観点を一つ明らかにしておく．それは，

(8) **言語は語の生態系** (ecological system of words) で，
(9) 一つのコーパスは (典型的には)，語の生態系の**時間切断面** (time-slice) である[4)]

ということである[5)].

以上の見解を発展させると，次のことが言える: **コーパス言語学とは，語を “種” と見なした場合の生態学，つまり “語の生態学” の実践に等しい**．これが意味することは，語 w の生起例 (occurrences/ instances) は生物個体 (individuals) に相同的であるということ，種の上に超種タクソンが存在しうること，こういうことがすべて，単なる比喩ではなく，ほぼ「文字通り」に該当し，それが「語の生態学」を可能にしている．

2.1.1 時間の止まった生態系

コーパスは (主に語の) 生態系であるが，それでも次の点には注意が必要である: **コーパス内部では時間が止まっている**[6)]．それは生態系の時間切断面だからである．

C は複数の種の個体群が構成する生態系の時間切断 (面) であるという観点が可能にする研究の幅は計り知れない．以下ではこの可能性について具体的に論じる．

2.1.2 「種」概念と「語」概念の相同性

コーパス C における語 w の挙動が生態的環境における生物種の挙動と相同だというのは議論を要する主張であるが，次の点に注目することが肝心である:

(10) **生物種** (species) **の概念と語の意味の概念は，変異の可能性も含めて，完全に一致する**[7)].

生物の生態系というものに関して十分な直観の働かない言語学者がこのことを理解するのは簡単ではないかも

---

4) 「典型的には」と断定を保留しているのは，コーパスは典型的には共時的な資料の集まりだが，中には通時的な調査のためのコーパス (Helsinki コーパスなど) が存在するからである．通時をもつコーパスは，成立場所を同じくする複数の時間断面の集まりである．

5) ただし，私はここで，語の概念をかなり大雑把に使用していることは断っておく．日本語で語の概念を定義することは —不可能ではないにせよ— 自明な課題ではない．

6) これは Web コーパスについては正しくない．だが，Web がコーパスと言えるかどうかには微妙な定義の問題もある．

7) 語の生態学としての言語記述と生物種の生態学とのアナロジーには一つ明らかな限界がある: 社会性生物の生態学は，言語の分析にはうまくあてはまらない．同種が群れている状態は，言語では同じ語が連続していることに相当するが，それは言語の場合には自然にありえることではない．

知れない．少し具体的に生態系がどんなものであるかを記述してみよう．

2.1.3 注意

言語が生態系であるというのは，確かに一つの喩えだが，**それは単なる喩え以上のものである**．これは説明を要する主張であり，以下ではその説明を試みる．

2.2 生態学とは?: インフォーマルな定義

まず生態学とは何であるかを簡単に解説しておこう．

ここで言う (生物の) **生態学** (ecology) とは生物個体 x (ないしは個体群 $X = \{ x[1], x[2], \ldots, x[n] \}$) と x (ないし X) の**環境** (environment) $E(x)$ (ないしは $E(X)$) の関係 $R(x,E(x))$ (ないしは $R(X,E(X))$) の研究のことである．R のことを**生態系** (ecological system) とも言う．

2.2.1 相互作用の科学

環境には他の生物個体 (同種の個体の集まり，異種の個体群の集まり) が含まれる．従って，生態学とは，個体 x とその環境 $E(x)$，個体群 X とその環境 $E(X)$ の**相互作用** (interaction) の科学である．

x も $E(x)$ も時間につれて変化するので，R も時間につれて変化する．R の時点 t における記述が，その時間切断面 $R(x(t),E(x(t)))$ である．

もう少し具体的に提唱の内容を言うと，それは

(11) コーパス C を一つの抽象的 “地理” (= 生態学的配置 (ecological geography)) だと見なすことであり，より具体的には，
   a. 語 w の生起位置 (position) $p(w)$ は，C という抽象的な空間内部の住み処 (locus) であり，
   b. 語 w の生起位置 $p(w)$ は，$p(w)$ の C 内での**生態的ニッチ** (ecological niche) を反映している，

   と考えることである．

すでに見たように，種 s の個体 $s[i]$ にとっての生態系とは $s[i]$ を取り囲む自然環境 (natural environment) $E(s[i])$ のことである．ここでは，種 s の個体の全体集合を $S = \{ s[1], s[2], \ldots, s[n] \}$ とすると，任意の個体 $s[i]$ にとって自然環境には，物理環境の他に (i) 他の種 $\bar{s}$ の個体の集合 $\bar{S}$，並びに (ii) 同種の他の個体 $S - s[i]$ がすべて含まれることに注意しよう．つまり，**個体 $s[i]$ にとっての生態環境とは，“自分” と “非自分” との関係の全体である**．この意味で，生態系とは個体について相対的なものである．

2.2.2 生態系の有限性と生態系の二つのクラス

生態系は，まず第一に個体にとっての環境である．第二に，それは種にとっての環境である．

第一の点に関しては，どんなに多く移動する種でも個体の活動範囲は有限なので，その意味で個体にとっての生態系は有限である．第二の点に関しては，個体群の行動範囲は有限なので，種にとっての生態系も有限である．ここには少し飛躍があるが，その効果が無視できるのは，次に示す個体群の定義から明らかであろう．

コーパスが語の生態系 (の時間断面) であるというアナロジーで役立つ側面は，前者の個体にとっての生態系の概念であり，後者の種にとって (正確には個体群にとって) の生態系の概念ではない．その意味ではコーパスが語の生態系の (時間断面) であるというアナロジーは不完全で，限界がある．

2.2.3 棲息条件という概念の抽象化と「抽象的生態学」

どんな生物種 (のどんな個体) にも様々な**棲息条件**が課せられる．嫌気性細菌以外の生物は酸素がない場所では生きられない．呼吸の条件は低次の，相対的に厳しい棲息条件の一つである．呼吸器の構造も大きな制約となる．エラ呼吸をする水棲生物種と肺呼吸をする陸棲生物種の差は大きい．

棲息条件にはもっと高次で，緩やかなものもある．例えば，温度や餌の条件がその例である．どんな生物種も適温を外れた地域にはいないし，どんな生物種もエサとなる資源のないところにはいない．適温の問題は，比較的低次の条件だが，摂食は，かなりの変異を許す．

語の生態系という喩えを有効にするため，棲息条件の多元性に拘らず，抽象的に (12) の考えで統一しよう:

(12) a. 種 s は生態系 S 内での存在のために，ある種の資源 r に依存する．
   b. r は S 内の至るところで利用可能ではない．
   c. このため，S 内の s の存在位置は r が存在する場所 (location) $L[+r]$ ($\in S$) に限定される

この考えは s が生命体であるか否かに係わらず成立する．このような特徴の研究を**抽象的生態 (学)** (abstract ecology) としよう．以下では，この抽象生態学の概念を基盤に語の生態系を研究するという方法を素描する．

2.3 抽象化された生態系の概念の帰結

この節では，言語を語の抽象的生態系だと見なすことによって得られる幾つかの簡単な帰結を示す．

2.3.1 個体としての語 w の棲息環境の記述

語 w が種であるとするならば，語の使用 (例) は個体であり，語の意味が種として同定される．語の使用例を $w[i]$ と表わそう．言語の場合，$w[i]$ の棲息環境は，$w[i]$ の生起環境 $E(w[i])$，具体的には $w[i]$ を含む語句 (phrase) であり，節 (clause) であり，文 (sentence) である．生態系の有限性によって，$w[i]$ の抽象的生態系には範囲がある — 正確にどこまでがそうなのを言うことは難しいけれど，それは不可能ではないだろう．

2.3.2 言語の単位は “語” ではなく “語の使用” である

まず抽象的生態系と見なすアプローチの本質的に重要な含意を明示化しよう．それは，

(13) **生態系としての言語の単位は語の使用 (uses) で**，
(14) **語の使用が個体に，語の用法 (usages) の確立が種分化 (speciations) に相当する**

ということである．

2.3.3 “用法” と “意義” の抽象生態学的定義

これは言語学の記述単位単位は語彙素 (lexemes) ではないということであり，これが言語が用法基盤であることの正確な意味である．従って，語 w (の意味) の **(使) 用法** (usages) とは，語 w (の意味) の種分化，変種化である．用法は**意義** (senses) とカップリングしている (はず).

2.3.4 “選択制限” の抽象生態学的定義

今日の技術をもってすれば，C について，$E(w[i])$ を網羅的に列挙することが容易である．そのための方法は KWIC である．x を語に限らないで，一般に語句とする．コーパス C を x について KWIC 検索すると，通常，数例から数百例 (コーパスの規模と検索語句の使用頻度によっては数万例) が見つかる．この検索は，x の棲息場所 $L(x)$ (= 生起環境 $E(x)$) を含んでいる．

$E(x) = L(x)$ は個体の棲息条件を満足している場所であるから，x が種であるならば，それは，種 x の棲息条件と，その**適応** (adaptation) とを同時に体現している．これが意味するのは，語句 x の示す**選択制限** (selectional restrictions) の特定とは，x の棲息条件と適応条件の特定にほかならないということである．極端な話をすれば，語句 x の示す選択制限とは x が “どんな食べ物，気候，地理” が好みか—そういう好みのことである．

この見地から得られる選択制限の理解では，選択制限は必ずしも主要部から項への一方的な要求ではない．共生関係にある語は互いに特定の種類の要素と共起することを選ぶ．それの一部が主要部だというだけの話である．このような選好の双方向性という形で語彙の生起条件に関して主流の言語理論で見られる主要部と項の関係の非対称性の仮定を弱めるという方向づけは，FOCAL [7] が体現する**並列分散意味論** (parallel distributed semantics) [6] の重要な知見であり，フレーム意味論/FrameNet [3] と洞察を共有する．

2.3.5 連語性の基盤

以上の洞察から得られるアナロジーの一つは，**ある種の個体の存在は，他の種の個体にとって資源 —正確には何らかの活動へのアフォーダンス [5, 8]— となる**ということである．**連語** (collocations) の基本的性質は，このアナロジーによって特徴づけが可能である．

2.3.6 “多義性” の抽象生態学的定義

語 w の**多義性**を記述し，特定するとは結局，w の種分化を記述し，特定することに等しい．それは w の (意味) 分化が —環境を変化させつつ— 環境へ適応し，隔離された $w[i]$ の個体群だと見なせるからである．

2.4 生物の生態学を言語学者が学ぶ必要性?

以上のアナロジーが妥当ならば，次のことが言える:

(15) コーパスを利用する言語学者は，(生物の) 生態学を学ぶことで恩恵を得ることができるし，そのような必要性は今後，どんどん大きくなってゆく．

実際，私は自分の経験から言語学者が (生物の) 生態学を学ぶことによって得られるものが非常に多いと強く確信している．それは**個体の特徴マーキング**に代表される観察・測定法，**多変量解析** (multivariate analysis) のようなデータ解析法のような具体的技能に限らず，「何を，どう見るか」に関する方法論的側面でも言えることである．

これが嬉しいことだと思うか，面倒で厄介なことだと思うかは，研究者の資質が問われる問題であろう．「手抜き」言語学に慣れた怠惰な人にとっては，これは少なからず苦痛なことだろうと容易に予想しうる．

3 おわりに

私はまず，コーパス言語学に関する (認知) 言語学内部での無理解を解消しようと思い，次のことを確認した: コーパス言語学を語彙分類学，その手段としての統計言語学と同一視するのは，(無知に基づく) 単なる短絡であり，かつ偏見である．コーパス言語学の可能性は，それより遥かに広く豊かである．それは，観測内容の質的向上の恩恵にあずかった研究の好例である．

この延長上にコーパスを利用する認知言語学が位置づけられるならば，それは (自然) 科学の標準的から見て量的にも質的にも不十分な従来の内観中心の研究法の限界を乗り越え，「言語学は言語の科学である」という内実のない定義の下にエセ科学がまかり通る時代の終焉を加速させると期待される．

参考文献

[1] M. Barlow and S. Kemmer, editors. *Usage-based Models of Language*. CSLI Publications, Stanford, CA, 2000.

[2] N. A. Chomsky. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA, 1965.

[3] C. J. Fillmore, C. R. Johnson, and M. R. L. Petruck. Background to FrameNet. *International Journal of Lexicography*, Vol. 16, No. 3, pp. 235–250, 2003.

[4] R. W. Langacker. A usage-based model. In B. Rudzka-Östyn, editor, *Topics in Cognitive Linguistics*, pp. 127–161. John Benjamins, Amsterdam/Philadelphia, 1988.

[5] E. S. Reed. *Encountering the World: Towards an Ecological Psychology*. Oxford University Press, 1996. [邦訳: 『アフォーダンスの心理学』. 細田直哉 (訳). 新曜社.].

[6] 黒田航, 井佐原均. 複層意味フレーム分析は言語表現を知識構造に結びつける: 文 “x が y を襲う” の理解を可能にする意味フレーム群の特定. In *KLS 25*, pp. 326–336, 2005.

[7] 黒田航, 中本敬子, 野澤元. 意味フレームに基づく概念分析の理論と実践. 山梨正明ほか（編）, 認知言語学論考第 4 巻. ひつじ書房, 133–269. [増補改訂版: `http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/roles-and-frames.pdf`].

[8] 佐々木正人. アフォーダンス: 新しい認知の理論. 岩波科学ライブラリー, 1994.