

用法基盤アプローチの理論的基礎*

コーパスの利用法に関する誤解を解く

Corpus Linguistics and Beyond

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

Modified 10/09/2007, Created on MM/DD/2005

1 はじめに

日本でもコーパス言語学 (Corpus Linguistics) [3, 5, 35] と呼ばれる分野が徐々に確立しつつあり、多くの研究者がその恩恵に預かるようになった。COBUILD をその嚆矢とするすぐれた辞書類は、コーパス言語学の賜物の一つである¹⁾。

だが、現状では、コーパス言語学が何であるか — とりわけそれが何の、どんな研究であるか — に関して、大きな誤解をしている人が多い。このような間違いをする人は、従来の「例文を作って勝負する」タイプの分析的手法な言語学で仕事をしてきた人に多い。この試論の目的は、その誤解を列挙し、その原因をあげ、それを通じてコーパス言語学の正しい理解 — 特にコーパスの利用価値に関して正しい理解 — を得ることを促進することである。

しかし、本論に入る前に、私がこのような文書を書くことになった背景を説明しておくことは有用であろう。特に私がなぜコーパス調査を強く推奨するのか、その理由を明確にしておくのは、本論の理解の助けになるだろう。

1.1 言語学者無用の時代?

言語学者は一時期、認知科学の花形であった。少なくとも MIT の偉大な言語学者の威光に効果のあった時代には、そうだった。だが、すでに凋落は始まっている。

機械学習 (Machine Learning) の進歩は凄まじい。その精度は平凡な言語学者が明らかにできる以上のことが明らかにできるほどになってきた。言語学者は工学者がコーパスから機械学習によって明らかできる以上のことを明らかにできなければ、もはや積極的な存在意義がない時代がすでに始まっている。

だが、コーパスから機械学習によって解ること以上のことができるためには、そもそもコーパスから何が解るのが解っていないければ、話にならない。言語学者は、はたして、この苦難の時代に生き残る技能を持っているのか?

私はこれを強く危惧する。これが私がコーパスの利用を強く訴える最大の理由である。

1.2 用法基盤主義に実質を与える必要性

私がこのノートを書こうと思ったもう一つの動機は、今や認知言語学から外部まで波及しつつある用法基盤=使用基盤アプローチ (usage-based approach) という重要な方法論上の概念の実質化が必要だと考えたからである。

1.2.1 用語の一人歩き

認知言語学の研究者は [4, 10, 24] を受けて (チョムスキー派生成文法の文法基盤~ 能力基盤アプローチ (grammar-based approach ~ competence-based approach) との差別化のために)、自分の研究が用法基盤=使用基盤アプローチであると宣言することが稀ではない。だが、奇妙なことに用法基盤アプローチを名乗る研究の多くは今だにチョム

* このノートの以前の題名は「コーパスの利用法に関する誤解を解く: コーパスを“時間の止まった生態系”と見なして“語の生態”を調査するために」だったが、10/09/2007の改訂に合わせて修正した。

¹⁾ ここで言う“コーパス” (corpus) とは、電子化された文書のことである。これは有効で、“事実上標準的”なコーパスの定義であるが、実は正確ではない。本来の意味でのコーパスは“時間的、空間的位置、使用(群)を同一にする言語使用の記録”のことであって、それは必ずしも電子化されている必要はない。また、コーパスは“アーカイブ” (archive) とは異なる。アーカイブは典型的には、幾つかの異なるコーパスを含む。この辺の事情に関しては、[3, 5, 35] のような入門書を参照のこと。

スキー派生成言語学と同じく作例中心の研究であり²⁾、コーパスの実地調査に基づいた実証的研究は少数派であるように思われる。

用法を体系的に収集し、観察しないで実践可能な用法基盤アプローチは—例えば「空気を含んだ真空」のように—完全に矛盾した概念である。実際、用法を調査せずに用法基盤主義を標榜することは、私には知的不誠実以外の何ものにも見えない³⁾。要するに用法基盤アプローチという用語だけが一人歩きして実質がない。

理念の明白な食い違いに由来する用語の一人歩きは、なるべく早期に解消される必要があると私は考える。だが、認知言語学に限らず多くの人の実感では「用法基盤の研究と言っても、いったい何をどうしたら良いかわからない」というのが正直なところではないだろうか？ 実際、「用法基盤アプローチ」という特定の名称を冠しているかどうかは別にして、用法の調査法に関する系統立った手引き書は、日本の言語学には存在しない⁴⁾。これでは確かに、どうしようもない。

1.2.2 時代の要請

これはとりもなおさず、日本の(少なくともチョムスキー革命以来の)言語学、生成言語学 vs 認知言語学/機能言語学の別を問わず、いわゆる理論言語学(theoretical linguistics)の実践体制では(説明に比べて)記述があまりに軽視されているということである。この体制を変革しない限り、どうしようもない。

コーパス言語学はチョムスキー派言語学に代表される極端な能力主義言語学への有力な対抗馬としてイギリスを中心に発達してきた[5, 15]が、これは私には機械学習の発達などを背景とした時代の要請であるように思われる⁵⁾。私はコーパス言語学の方

法論の認知言語学への積極的な導入がその変革に繋がると考え、そのための方法論的、理論的基礎を敷くことを試みることにした。

1.3 議論の構成

以上の動機の下、§2 で典型的な誤解を解くことを試みる。その延長上に §3 でコーパスと生態系とのアナロジーを展開する。§4 で暫定的な結論を述べる。§ 付録 A ではコーパス言語学を支えて来た技術の幾つかを紹介する。

2 コーパス言語学に関する幾つかの誤解を解く

2.1 コーパス言語学は何でないか？

議論に先立って、次のことは確認しておく必要がある：

- (1) コーパスの利用法はただ一つしかないわけではない(実際、コーパスの本当の利用価値は、その多目的性にある)。
- (2) 特に、語句の数(頻度)を数え、それに基づいて語彙を分類することは唯一のコーパスの利用法ではない。

語句の数(頻度)を数え、それに基づいて語彙を分類すること—それは確かに、コーパスの典型的な利用法(の一つ)であり、この利用法の上に成立している研究は数も多く、また大きな成果を上げている。だが、(1)に明示したように、コーパスの利用法はこれに限られるわけではない。

問題なのは、語句の語句の数(頻度)を数え、それに基づいて語彙を分類する仕事にコーパス言語学全体を同一視している研究者は、言語学の内外にも驚くほど多数いるということである。そういう人たちはコーパスの可能性について—従って、コーパス言語学の可能性について—根本的に誤解している。その理由は以下に示す。

2.1.1 誤解の内容

コーパスの利用価値に関して、上で説明したような誤解をしている人は典型的には、次の症例を呈する：

- (3) いわゆるコーパス言語学の名の下に公表される研究成果(これを“狭義の”コーパス言語学と

理論の成果の一つである。

²⁾ これは R. Langacker の研究(のほぼ全体)も含まれる。

³⁾ 私が思うに、用法基盤アプローチは今のところ、能力基盤アプローチの対抗馬として、認知言語学者の「公約」として掲げられているにすぎない。だが、認知言語学内部でも、例えば認知意味論[23]が本当に用法基盤アプローチと互換性があるか不明である[34]など、ハッキリしない点が多い。

⁴⁾ その理由は明らかに日本の多くの言語学の研究室では用例の収集法、分析法に関する指導が—少なくともチョムスキー派生成言語学の度を外れた受容以来—まったくと言って良いほど行われて来なかったからである。

⁵⁾ 些か皮肉なことだが、機会学習の前提となる構文パーサー(syntactic parsers)はチョムスキーの初期の計算言語学

する)を、コーパスを対象にした調査から得られる研究成果(これを“広義の”コーパス言語学、あるいはコーパス言語学の可能性(Corpus Linguistics Potentials)とする)が区別できない。そのため、

- (4) a. コーパス言語学を統計/計量/数理言語学(Statistical Linguistics, Mathematical Linguistics)⁶⁾と混同している。
 b. コーパス言語学を(定量的)語彙分類(Lexical Classification/Linguistic Taxonomy)と混同している。
 c. コーパス言語学を辞書編纂学(Lexicography)と混同している。

(4)の誤解はいずれも(3)から派生する結果である。以下では、(4a, b, c)の誤解について、少し詳しく説明する。

2.1.2 コーパス言語学は統計/計量言語学ではない

言語学に限らず、統計は役に立つ。だが、統計は手段であって、目的ではない。統計を利用するという手段の有効性は、目的の有効性によって決まる。従って、統計がなにゆえ必要とされるのかが理解されない限り、統計の意義は理解されない。

統計/計量/数理言語学と呼ばれる多く研究[33, 37]は、比較的抽象的な語彙の特性に焦点を当て、すぐれた成果を上げている。

だが、ここで理解しておく必要のあるのは、次のことである: 従来の統計/計量言語学の研究成果が示しているのは、語彙特性を調べるといった目的のためにコーパスの利用が有効だということではない。それ以上の意味はない。コーパスの利用可能性は、ほかにまだまだある。広義のコーパス言語学の研究対象は、この統計情報に限られるわけではないし、限られるべきではない。

結論: コーパス言語学を統計/計量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。

次に残りの誤解(4b, c)について考察することにする。

2.1.3 コーパス言語学は語彙分類学ではない

(4b, c)の誤解は互いに関連している。というのは、語彙分類は辞書編纂の基礎研究をなすからである。Collins COBUILDが“Bank of English”[18]と呼ばれる大規模なコーパスを基礎にして画期的な成果を挙げたのは、記憶に新しい。同様の方法論で、日本でも井上・赤野が[32]が画期的な辞書を編纂した。コーパスに拠る所を求める文法記述の枠組みには、Pattern Grammar[17], Data Oriented Parsing[6, 7]のような試みもある。

コーパスの利用は確かに、辞書の編纂に役に立つ。有用な辞書は、統計にしか現れない、微妙な語彙の特性を反映する語の定義を示す必要があり、コーパスの利用は、その目的のために実際に有効である。コーパスの利用が有効である事情は、統計/計量言語学の場合と似ている。

だが、ここで改めて理解しておく必要のあるのは、(4a)の場合と同様、次のことである: 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った目的のためにコーパスの利用が有効だということではない。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。

結論: 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。

これは今となっては言うまでもなく明らかなことであろう。

2.1.4 簡単なまとめ

以上、コーパス言語学に関する誤解を解くために、主にそれが何でないかを見たきた。以下ではコーパス言語学がそれ以外の何でありうるか、つまりコーパス言語学の可能性に関して、もう少し詳しく考察するつもりである。だが、その前に、一つ確認しておきたいことがある。

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。世の中、分類が趣味の人々がいたって、

⁶⁾ 日本で mathematical linguistics と呼ばれているものと欧米で同一の名称で呼ばれているものは、ずいぶん内容が違う。Z. Harrisの仕事(例えば[16])などが欧米の mathematical linguistics の典型であるようだ。

ぜんぜんおかしくないのだ。それは正しい趣味であり、ほかの人にとって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に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。

一つの重要なのは、コーパス言語学者自身が語彙分類の上に何が可能なかが解っていない状態にある、という可能性である。そうだとすれば、今のコーパス言語学は、喩えて言うなら、進化論のなかった時代の生物学のような状態だ: Darwin 以前、生物学は分類以外に何もすることはなかった。これは喩えだが、それでも私はこれが現在のコーパス言語学の適切な評価であるように思う。コーパス言語学には早く Darwin が現われ、彼と共に次の段階に進むべきなのである。

以下、私はそのための歩みを促進させるために一つのアナロジーを展開する。その際、主導的な役割を演じる主題は、次のものである: では、なぜコーパスはこのような目的のための有用なのか? — この問いには誰も思いを駆せていないように私には思われる。この問いの答えを私は部分的に §3.2 で提供したいと思うが、本論に入る前に少し方法論的な下準備をしよう。

3 用法基盤言語学への道: 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。この節ではそれを明らかにすることを試みる。

3.1 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、それが言語の用法基盤のアプローチ (usage-based approach to language) の実践に最適なモデル化であると主張する。

3.2 コーパスは語の生態系の時間断面である

私がコーパスの有意義な利用法を理解し、開発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、

- (7) 言語は語の生態系 (ecological system of words) であり、
- (8) 一つのコーパスは (典型的には)、その生態系の時間断面 (time-slice) である⁷⁾

ということである⁸⁾。

3.2.1 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。もち

⁷⁾ 「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、中には通時的な資料からなるコーパス (Helsinki コーパス [22] など) が存在するからである。通時をもつコーパスは複数の時間断面の集まりである。

⁸⁾ 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。

ろん, これは説明を要する主張であり, 以下で私はその説明を試みる.

3.2.2 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 $X = \{x[1], x[2], \dots, x[n]\}$) 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.

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

- (9) 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,
- ある語 w の生起位置(position) $p(w)$ は, C という抽象的な空間内部の“場所”(location)であり,
 - ある語 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.

以下, この考えを

3.2.3 複雑系としての言語, 生態系としての言語

言語は単に複雑系(complex system) [31] であるばかりではない. それは生態系(ecological system), 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, 生態系における種概念と言語における語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.

生態系における種の定義に関する考察は E. Reed の論考 [27, 28] を参照した.

3.2.4 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⁹⁾. それは生態系の時間切断面だからである.

これには良い面と悪い面がある: 言い面としては, 時間発展以外の静的構造はうまく保存されていて, 十分に豊かである. この点に関しては, §3.6 でもう少し詳しく論じることになろう.

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, 具体的に論じることにする.

3.3 “語”の生態学

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, 文字通り該当し, それが語の生態学を可能にしている.

コーパス C における語 w のふるまいとは, 生態的環境における生物種のふるまい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: 生物種(species)の概念と語の概念は, 変異の可能性も含めて, 完全に一致する¹⁰⁾.

生物の生態系というものに関して十分な直観の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知れない. 少し具体的に生態系がどんなもので

⁹⁾ ただし, これは Web コーパスについては正しくない. しかし, そもそも 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.

¹⁰⁾ 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.

あるかを記述してみよう。

3.3.1 生態系とは?

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 $S = \{s[1], s[2], \dots, s[n]\}$ 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。

3.3.2 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。

3.3.3 個体群の発生と生殖隔離

繁殖行動の際、お互いの行動範囲内に相手がいないければ、個体は相手と接触することができず、結果として、繁殖行動ができない。これは個体の集合を一定に空間的範囲内に閉じこめる効果をもつ。このような閉じこめによって生じる生物の生活単位は個体群 (population) と呼ばれる。

個体群は空間的に局所化され、互いに隔離される傾向がある。これを生殖隔離 (reproductive isolation) と言う。個体群が進化の単位となるのは、この生殖隔離の効果である。ただ、実際の生態学の場合と異なり、語の生態学では生殖に関する問題は起こらないので、この生殖隔離に関するアナロジーはコーパス言語学には成立しない。

3.3.4 生態系の概念の抽象化

どんな生物種のどんな個体にも棲息条件がある。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。これは低次の、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。

棲息条件はもっと高次で、緩やかなものもある。例えば、どんな生物種も適温を外れた地域にはいない。また、どんな生物種もエサとなる資源のいないところにはいない。適温の問題は、比較的 low 条件だが、どんなものをエサとするかは、かなりの変異がある。

非常に多くの生物種が他の生物種をエサにしている。実際、無生物をエサにする生物種の幅は小さい。知られている限りで、植物以外にそのような生物種は存在しないようだ。

ただ、語の生態系というアナロジーを有効にするため、ここでは摂食という側面にこだわらず、抽象的に (10) の考えで統一しよう。

- (10) 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して、語の生態系を研究するという方法を素描しようと思う。

3.4 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。

3.4.1 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、生態系としての言語の単位は語の使用 (例) (uses) である、という

ことである¹¹⁾。これは言語学の記述単位単位は語ではなくて、語の使用だということであり、これが言語が使用=用法基盤 (usage-based) であることの正確な意味である。

3.4.2 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) は個体である。使用を $w[i]$ 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のような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。

生態学の非常に重要な洞察の一つは、ある種の個体の存在は、他の種の個体にとって資源 — 正確にはアフォーダンス (affordances) [28, 36] — となるということである。

生態学的見地から得られる選択制限の理解によれば、選択制限は必ずしも主要部から頂への一方的な要求ではない、ということである。語は互いに特定の種類の要素と共起することを選ぶ。その一部が主要部だというだけの話である。主要部/項の非対称性を弱めるという方向づけは、分散意味論 (distributed semantics) の重要な知見である。

3.4.3 “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 KWIC¹²⁾ 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、ある語句 x の示す選択制限とは x が“何を食べ”て、“どこで寝る”のが好

きか—そういう好みのことである。

3.4.4 “多義性”の抽象生態学的定義

語 w の多義性 (polysemy) を記述し、特定することは結局、 w の種分化 (speciation) 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。

3.4.5 “用法”と“意義”の抽象生態学的定義

語 w (の意味) の用法 (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)。

3.5 生物の生態学を言語学者が学ぶ必要性

以上のことが正しければ、次のことが言える。

- (11) a. コーパスを利用する言語学者は、(生物の) 生態学の基礎を学ぶことで非常に大きな恩恵を得ることができる。
- b. 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確信している。それは個体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。

3.5.1 とりあえず生態学の入門書を幾つか

生態学の入門書は、例えば次のようなものがある:

- 『トマトはなぜ赤い — 生態学入門』。三島次郎。東洋出版社。1992。

生物学の知識をまったく前提としない、わかりやすい生態学の入門書。説明は懇切丁寧、数式は最低限、更に、近接要因 (proximate factors) と究極要因 (ultimate factors) の区別の説明 (p. 13) など、人文系の読者にも本当に親切かつ、有益だと思われる情報を提供してくれる

¹¹⁾ 意味素性が表現型 — あるいは遺伝子 — に相当するだろう

¹²⁾ この略称は、“Key Word In Context” から来ている。

本だと思う。「自然界におけるレベル」(p. 7)をハッキリ明示し、「全体から部分へ」(p. 8)の説明の必要性を訴えるところなど、複雑系の思考の本質を捉えた説明が冴える。非常にオススメ度の高い本。

- 『生態と環境 — 生物学入門コース(7)』. 松本忠夫. 岩波書店. 1993.

どちらかという生態環境の記述というより、適応戦略として見た種の行動の記述が多い。記述は平易で、わかりやすい。ついでに言うと、手書きの図がキレイでカワイイ¹³⁾。おそらく生物学の知識のない人文系の読者にでも十分理解できる内容だろう。ただ、細胞学、分子生物学、遺伝子学のようなミクロ生物学をの除いて、行動学を中心に組み立てたマクロ生物学への入門書と言った感じで、生態学の観点からすると、もの足りない。

- 『ゼロからわかる生態学 — 環境・進化・持続可能性の科学』. 松田裕之. 共立出版. 2004.

数理モデルによる個体群生物学 — 特に個体群間のダイナミクス — に重点がある。個々の個体(群)にとっての棲息環境の記述法の問題に関しては、あまり詳しくない。生物学の知識のない人文系の読者に簡単に理解できる内容とは思われない(生物学を専門としない理工系の読者がターゲットだと思う)ので、超初心者へのオススメ度は低いが、数理モデル化に興味野ある人には勧められる。

このような解説書を有意義に読む秘訣は、書いてあることを文字通りに取らないことである: 常にアナロジーを働かせ、種を語に、個体を語の生起例に置き換えて読むと、理解が進むだろう。ただ、私は生態学の専門家ではないので、これらの情報源の専門家から見た有用性は保証できない。

3.5.2 最後に統計の話を再び

私は、§2.1.2で「言語学に限らず、統計は役に立つ」と言った。実際、統計は本当に役に立つ。統計が役に立たない分野は自然科学の中には存在しないほどである。

統計の重要性は生態学に関して特に言えることで

ある。生態学では伝統的に統計解析の比重が強い。生態を調べるには、環境自体の特徴を直接観測、計測することは難しく、その反映である個体群の個体数、個体群の平均体長、平均重量、平均寿命などの抽象的性質の測定が、しばしば色や形などの具体的な形質の観測より(相対的に)重要な意味をもってくるからである。

これは言語学に通じる話である。これまでコーパス言語学が統計を利用して明らかにしてきた性質は、実際、こういう抽象的な性質であることが多かった。

だが、統計の対象を抽象的な性質に限るのは必然的なことではない。色、形のような具体的な特徴も統計的解析手法で捉えることができる。厄介なのは、これれが体長、体重、個体数のような特徴より計測しにくいという点にある。このような複雑な形質(characteristics)を統計的に解析しようとすれば、個体マーキングのような操作が必要となる。

コーパスを通じて、言語の何を見ようと思うかは、「どうすれば何がどうみえるのか」に関する知識に依存する。生態学的アナロジーは、この点で言語学者の想像力を大いに拡大してくれるはずである。この種の手法の導入による、記述可能な対象の拡大が、コーパス言語学の可能性を広げると私は考える。

3.6 補足: コーパスの無時間性の意味

すでに注意を促しておいたように、コーパス内部の時間は止められている。それには良い面も悪い面もある。

3.6.1 無時間性の嬉しくない面

コーパスの時間が(通常は)固定されていることの嬉しくない面は、個体の発生、種の進化に相当する現象が見れないという点である。

ただ、解決は意外と簡単である。語の(意味構造の、音韻構造の)の進化を調べたい場合には同一個体群について異なる時間断面をトレースできるようなコーパス利用法を考えれば良い。実際、Helsinki Corpus [22]のように、そのような言語進化の記述のために開発されたコーパスも存在し、少なくとも英語に関して、限定的にはあるけれど、語と文法の進化を記述する可能性を提供している。

¹³⁾ 生物嫌いの多い女性学生に嫌われないように、女性の書いた挿絵が採用されているのだと思われる

3.6.2 無時間性の嬉しい面

コーパス内部で時間が進行しないということが意味する良い面の一つは、コーパスの中の個体は不死だということである¹⁴⁾。別の言い方をすれば、何をしてもよい。ここで少し考えて欲しい。どんなにぞんざいに扱っても個体が死なないというのは、進化や個体発生が研究の目的ではない限り、理想的な実験環境である。ちょっとした失敗で飼育生物群を殺してしまう可能性と日々闘っている実験生物学者には羨ましい限りだろう。

何をしてもよいというのは強烈な特権なのである。なぜもっと多くの人がこの特権を利用しないのだろうか?と、私はいつも不思議でならない。

4 おわりに

この試論で私はコーパス言語学に関する言語学内部での誤解を解こうと思った。そのために、次のことを確認した:

- コーパス言語学を統計/計量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。
- それと同じく、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。

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、言語資源の充実による、言語を対象にする観測可能性の増大の恩恵にあずかる研究の代表例である。

この延長上にコーパス言語学が位置づけられるならば、それは、(自然)科学の標準的レベルから見て質的も量的も不十分な内観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。

付録 A コーパス言語学を支える基礎技術

言語学内部では(周辺分野への無知故に)正当に評価されていないが、コーパス言語学の発展を支えてきた重要な基礎技術群が存在する。それは形態素

解析(morphological analysis)、品詞タグづけ(part of speech tagging)、統語解析(syntactic parsing)であり、その産物が解析済みコーパス(parsed corpus/corpora)、注釈つきコーパス(annotated corpus/corpora)である(corpusの複数形はcorpora)。解析コーパスの代表例は、Penn TreeBanks [1]、京大コーパス [38, 21] で、(統語解析なしの)注釈つきコーパスの代表例は、Brown Corpus [13]、British National Corpus (BNC) [9] のような品詞タグつきコーパスである。この意味で、タグづけは、注釈(annotation)の一種と理解される [15]。注釈つきコーパス、解析済みコーパスは共に言語資源(language/linguistic resources)と呼ばれる。

解析のための枠組みで主要なものは、以下の通り:¹⁵⁾

- 句構造文法(Phrase Structure Grammar)タイプ: Head-Driven Phrase Structure Grammar (HPSG) [26]、Lexical Functional Grammar (LFG) [8]、(Lexicalized) Tree Adjoining Grammar ((L)TAG) [2]
- 依存文法(Dependency Grammar)タイプ: 係り受け解析(KNP) [38, 21]、Link Grammar [30]、Word Grammar [14, 29])

このような言語資源の開発の動機は、(特に日本では)工学的な背景から来るものであることが多く、必ずしも純粋に言語学的なものではないが、それでも言語学は、十分にこのような言語資源の恩恵にあずかっていると思う。このことに関する正しい理解は、コーパス言語学の可能性を正しく見極めるためにも不可欠である。この意味で、言語資源の開発のための基本概念を見ておくことは、有意義であると私は考える。この節はその目的のために書かれた。

A.1 タグつきコーパスの価値: 言語観察の質、量の向上

言語学から見た言語資源の恩恵を一言で言うならば、それは言語を対象にする観測可能性の増大であ

¹⁴⁾ これは人工生命(Artificial Life) [25] の場合と状況が似ている。

¹⁵⁾ この一覧を見ても判る通り、統語解析に関心をもっているのは、一部の言語理論に限られる。Word Grammar [14, 29] を除いて認知言語学の諸理論は、それに含まれていない。これは認知言語学が人文系の「実用性無視」の伝統を強く引きずっていることの証であろうか?

る。どんなにすぐれた洞察力と記憶力をもった優秀な言語研究者であっても、一人の人間が、紙と鉛筆と達成できる研究内容にはおのずから限界がある。それは、ヒトの肉眼では微生物が見えない、遠方にある天体の細部が見えないのと同じ観測上の限界である。

微生物学、生化学にとって顕微鏡が、天文学にとって望遠鏡が、それぞれの分野で新しい観測可能性を提供し、研究内容を拡充したように、言語学にとってタグつきコーパスは、新しい観測可能性を提供し、研究内容を拡充することになる。従って、注釈つきコーパスの役割は言語研究に決定的な役割を演じ、その利用は今後、不可欠なものとなってゆくだろう。将来的には、研究目的のために適切なコーパスを用いない研究には意義が認められない時代が来るだろう。この時代に至って、現在まで支配的だった「手抜き」言語学は終焉し、言語学はようやく科学的になる。このような時代の流れは不可避的であり、また望ましいことである、と私は信じる。

言語研究はすでに言語資源から大きな恩恵を受けているが、将来的には、この技術の延長上に意味タグつきコーパス、意味役割タグつきコーパス、談話タグつきコーパスが開発されることで、言語学の研究は飛躍的に発展する可能性がある。そのことを理解するためにも、タグづけの基礎を理解しておくことは極めて重要である。

現在、言語学内部では内観に基づく研究方法が支配的だが、それは — 少なくとも (自然) 科学の標準的レベルから見ると — 質的にも量的にもまったく不十分である。これは現在の言語研究の実態を見れば、明白である。言語資源の充実は、この種の科学的蛮行が「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まかり通る時代の終焉を加速させると期待される。

A.2 品詞タグつきコーパス

この節では品詞タグつきコーパスの説明をする。まず品詞タグづけ (“part of speech” tagging またの名を POS tagging) の定義から始めよう:

(12) 品詞タグづけとは、

- a. 品詞タグ体系 (pos tag system) $T = \{t_1, t_2, \dots, t_N\}$ が与えられ、

- b. コーパス C について、 C を構成する文の集合を $S(N) = \{s_1, s_2, \dots, s_N\}$ とするとき、
 c. おのおのの s について、その形態素解析列 $A(s) = [m_1, m_2, \dots, m_n]$ に対し、 T の要素の一つ t を割り当てる作業で、
 d. その結果は、 $l = [[m_1, t_i], [m_2, t_j], \dots, [m_n, t_k]]$ である (t_i, t_j は同じであっても構わない)。ただし、 $[m, t]$ は形態素 m と品詞タグ t の対である。

タグ体系は、タグ集合 (tagset) と呼ばれる。 T の具体例は、{ 普通名詞, 固有名詞, ..., 格助詞, 係助詞, ..., サ変他動詞, 助動詞, ... } のような品詞の集合である¹⁶⁾。

A.2.1 品詞タグづけ例 1

品詞タグづけの例を一つあげると、(13a) とし、その形態素解析列 $A(s)$ を (13b) とすると、品詞タグづけの一例は (13c) となる。

- (13) a. $s =$ “ある男が奈々子に失恋した”
 b. $A(s) =$ [“ある”, “男”, “が”, “奈々子”, “に”, “失恋し”, “た”]
 c. $T(s) =$ [[“ある”, 限定詞], [“男”, 普通名詞], [“が”, 格助詞], [“奈々子”, 固有名詞], [“に”, 格助詞], [“失恋し”, に格支配の他動詞], [“た”, 動詞活用語尾]]

品詞タグつきコーパス C というのは、 C を構成する (原理的には) すべての文について、このような品詞タグがついているもののことを言う。

A.2.2 品詞タグづけ例 2

(14) は、*Spoken English Corpus* の文 (15) を (16) に定義されている $C7$ タグセットでタグづけしたものである¹⁷⁾。

- (14) `Perdita&NN1-NP0; ,&PUN; covering&VVG; the&AT0; bottom&NN1; of&PRF; the&AT0; lorries&NN2; with&PRP; straw&NN1;`

¹⁶⁾ ほとんどの (半) 自動の品詞タグづけシステム (e.g., CLAWS2) で、システムが採用するタグ体系に含まれる品詞の総数は、それが標的としている言語の伝統的文法が認める分類よりも多い。典型的には $n \times (10 \sim 10^2)$ 個ぐらいの品詞の別がある。ただし、これは派生品詞を勘定に入れての話であり、基本品詞の数はもっと少ない。

¹⁷⁾ この節の解説は <http://bowland-files.lancs.ac.uk/monkey/ihe/linguistics/corpus2/2fra1.htm> に基づくものである。

to&TO0; protect&VVI; the&AT0;
ponies&NN2; '&POS; feet&NN2; ,&PUN;
suddenly&AV0; heard&VVD-VVN;
Alejandro&NN1-NP0; shouting&VVG;
that&CJT; she&PNP; better&AV0;
dig&VVB; out&AVP; a&AT0; pair&NN0;
of&PRF; clean&AJ0; breeches&NN2;
and&CJC; polish&VVB; her&DPS;
boots&NN2; ,&PUN; as&CJS; she&PNP;
'd&VM0; be&VBI; playing&VVG; in&PRP;
the&AT0; match&NN1; that&DT0;
afternoon&NN1; .&PUN;

- (15) Perdita, covering the bottom of the lorries with straw to protect the ponies' feet, suddenly heard Alejandro shouting that she better dig out a pair of clean breeches and polish her boots, as she'd be playing in the match that afternoon.
- (16) 1. AJ0: general adjective
2. AT0: article, neutral for number
3. AV0: general adverb
4. AVP: prepositional adverb
5. CJC: co-ordinating conjunction
6. CJS: subordinating conjunction
7. CJT: that conjunction
8. DPS: possessive determiner
9. DT0: singular determiner
10. NN0: common noun, neutral for number
11. NN1: singular common noun
12. NN2: plural common noun
13. NP0: proper noun
14. POS: genitive marker
15. PNP: pronoun
16. PRF: *of*
17. PRP: preposition
18. PUN: punctuation
19. TO0: infinitive *to*
20. VBI: *be*
21. VM0: modal auxiliary
22. VVB: base form of lexical verb
23. VVD: past tense form of lexical verb
24. VVG: *-ing* form of lexical verb
25. VVI: infinitive form of lexical verb
26. VVN: past participle form of lexical verb

重要な点は:

- (17) a. どのタグも三つ文字 (XXX) からなる .
b. タグは TEI (Text Encoding Initiative)¹⁸⁾ entity references の境界確定子 & and ; で語に付加されている .

一部の *heard* のような語には二つのタグが付与されている . この種のタグは、合併タグ (**portmanteau tags**) として知られ、コーパスが修正済みでない場合に使用者にタグづけ誤りである可能性を示唆するために付与されている .

A.2.3 品詞タグつきコーパスの効用

品詞タグつきコーパスの最大の恩恵の一つが、語彙情報検索の精度と能率が飛躍的に向上するという点である . 例えば、十分に強力な検索エンジンがあれば、「他動詞 *find* の目的語句 (object NP) として使われている普通名詞をすべて列挙せよ」という命令を実行し、その一覧を得ることができる . 同じ理屈で、他動詞 *find*, *look for* の目的語句として使われている普通名詞の一覧をおのおの N1, N2 とし、それらの共通項 ($N3 = N1 \cap N2$) と独自項 ($N1 - N3, N3 - N3$) を列挙する、ということも可能である . これは定量分析と定性分析に橋渡す . その成果の一つが辞書編纂である .

A.2.4 注意

$T(s)$ の項は、 $A(s), T$ である . 別の言い方をすれば、 $P(s)$ は $A(s)$ の結果と T にある定義に依存する . 一方、 T は $A(s)$ は相互依存する . このことから、 T を開発する作業は自明でないことがわかるだろう .

A.2.5 タグ体系の条件

品詞タグ体系に限らず、タグ体系 T が一般的に満たすべき条件として、網羅性、一貫性、完全性がある . 具体的には、次のようなものである:

- (18) コーパス C の文の集合 $S(N) = \{s_1, \dots, s_N\}$ に付けるタグ体系 T について、 $A(s) = [m_1, \dots, m_n]$ とするとき、
- 網羅性 (exhaustiveness):
すべての $A(s)$ について、少なくとも一つタグづけ $T(s)$ が存在し、
 - 完全性 (completeness):
かつ、そのタグづけが唯一の可能性であ

¹⁸⁾ <http://www.tei-c.org/P4X/>

る (つまり, $A(s_i) = A(s_j)$ ならば, $T(s_i) = T(s_j)$ である) .

- c. 一般性 (generality/applicability):
少なくとも一つ以上の C について以上の二つの条件が満たされる
- d. タグづけが完全性の条件が満たすならば, それに利用されたタグ体系 T は一貫性 (consistency) をもつ .

A.3 文法役割タグつきコーパス

品詞タグづけの概念を拡張して, 文法役割タグづけというものを定義することが可能である . 拡張のために必要なのは, T に条件を満足する文法役割を定義することである . だが, これはもちろん, 自明な課題ではない .

これを理解するには, T が $A(s)$ と T に依存すること, $A(s)$ と T が相互依存の関係にあることを思い出すだけで十分であろう . s の文法役割についてタグ体系 T_G を定義するには, T_G に見合った文法役割解析 $A_G(s)$ が定義されていなくてはならない . 形態素解析の場合は多かれ少なかれ一次元列で十分であったが, 形態素解析の場合と異なり, $A_G(s)$ が一次元列ではないのは明らかである . それは少なくとも木以上に複雑な構造である .

このような問題があるとは言え, Penn Treebank の Phase2 の結果を見ればわかるように, 少なくとも英語に関しては, 文法役割タグづけは部分的に成功している . 日本語に関しては, 係り受け解析 – これは, 依存構造 (dependency parsing) の一種である – の研究, 開発が進んでいるとは言え, 文法役割解析, タグづけはまだ発展途上の段階にある .

A.4 意味 (役割) タグつきコーパス

現在, 世界各地で複数の研究グループが, 意味タグつきコーパスの開発に着手している . 意味タグ体系に WordNet [11] を採用する意味タグつきコーパスの開発が有望視されている .

これとは別に意味役割をタグ体系に採用するという方向も存在する . FrameNet [12], PropBank [19, 20] がこの路線で進められている研究の代表例である .

A.5 統語解析済みコーパス, あるいは TreeBanks
統語解析済みコーパスの解説は本論文の内容とは係わりが薄いので, 割愛する . [1] のような図書を参考にされたい .

参考文献

- [1] A. Abeillé, editor. *TreeBanks: Building and Using Parsed Corpora*. Kluwer Academic Press, 2003.
- [2] A. Abeillé and O. Rambow, editors. *Tree Adjoining Grammars: Formalisms, Linguistic Analyses, and Processing*. CSLI Publications, Stanford, CA, 2000.
- [3] K. Aijmer and B. Altenberg. *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, 1991.
- [4] M. Barlow and S. Kemmer, editors. *2000*. CSLI Publications, 2000.
- [5] D. Biber, R. Reppen, and S. Conrad. *Corpus Linguistics*. Cambridge University Press, 1998. [邦訳: コーパス言語学: 言語構造と用法の研究. 齋藤ほか (訳). 南雲堂. 2003].
- [6] R. Bod. *Beyond Grammar: An Experience-Based Theory of Language*. CSLI Publications, 1998.
- [7] R. Bod. Extracting stochastic grammars from TreeBanks. In A. Abeillé and O. Rambow, editors, *TreeBanks*, pp. 333–349. Kluwer, 2003.
- [8] J. Bresnan. *Lexical-Functional Syntax*. Blackwell, 2001.
- [9] L. Burnard and G. Ashton. *The BNC Handbook: Exploring the British National Corpus with Sara*. Edinburgh University Press, 1998. [邦訳: コーパス言語学への誘い. 北村 裕 (訳). 松柏社.].
- [10] J. L. Bybee. *Phonology and Language Use*. Cambridge University Press, Cambridge, 2001.
- [11] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [12] T. Fontenelle, editor. *FrameNet and Frame Semantics*. Oxford University Press, 2003. A Special Issue of *International Journal of Lexicography*, 16 (3).
- [13] W. N. Francis and H. Kučera. *Brown Corpus Manual*. 1964-79.
- [14] N. Fraser. *Dependency Parsing*. Unpublished ph.d. dissertation, University College London, 1993.
- [15] R. Garside, G. Leech, and A. McEnery. *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman, 1997.
- [16] Z. S. Harris. *A Theory of Language and Information*. Oxford University Press, 1991.
- [17] S. Hunston and G. Francis. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins, 2000.
- [18] T. Järvinen. Bank of english and beyond. In *Tree-*

- Banks: Building and Using Parsed Corpora*, pp. 43–59. Kluwer Academic Press, 2003.
- [19] P. Kingsbury and M. Palmer. From TreeBank to PropBank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, 2002.
- [20] P. Kingsbury, M. Palmer, and M. Marcus. Adding semantic annotation to the Penn TreeBank. In *Proceedings of the Human Language Technology Conference*, 2002.
- [21] S. Kurohashi and M. Nagao. Building a Japanese parsed corpus: While improving the parsing system. In A. Abeillé, editor, *TreeBanks: Building and Using Parsed Corpora*, pp. 249–260. Kluwer Academic Publishers, Dordrecht/Boston/London, 2003.
- [22] M. Kytö. *Helsinki Corpus Manual*, 3rd edition edition, 1996. <http://khnt.hit.uib.no/icame/manuals/HC/>.
- [23] G. Lakoff. *Women, Fire, and Dangerous Things*. University of Chicago Press, 1987. [邦訳: 『認知意味論』(池上 嘉彦・河上 誓作 訳). 紀伊国屋書店.]
- [24] R. W. Langacker. A usage-based model. In B. Rudzka-Östyn, editor, *Topics in Cognitive Linguistics*, pp. 127–161. John Benjamins, Amsterdam/Philadelphia, 1988.
- [25] C. Langton, editor. *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems Held September, 1987 in Los Alamos*. Santa Fe Institute, 1989.
- [26] C. J. Pollard and I. A. Sag. *Head-driven Phrase Structure Grammar*. Studies in Contemporary Linguistics. Center for the Study of Language and Information/The University of Chicago Press, Stanford, CA/Chicago, IL, 1994.
- [27] E. S. Reed. Darwin’s evolutionary philosophy: The laws of change. *Acta Biotheoretica*, Vol. 27, No. 3/4, 1978. [邦訳: 『ダーウィン進化論の哲学. 細田直哉訳. 『アフォーダンスの構想』, 217–66. 佐々木 正人・三島 博之(編訳). 東京大学出版会.]
- [28] 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.]
- [29] O. Shaumyan. Parsing english with word grammar, 1995. Unpublished Master’s Thesis.
- [30] D. Sleator and Temperley, D. Parsing english with link grammar. In *Third International Workshop on Parsing Technologies*, 1993.
- [31] M. M. Waldrop. *Complexity*. Simon & Shuster, 1994. [邦訳: 『複雑系—科学革命の震源地・サンタフェ研究所の天才たち』(新潮文庫). M. ミッチェル・ワールドロップ(著). 田中 三彦・遠山 峻征(翻訳)].
- [32] 井上永幸, 赤野一郎. ウィズダム英和辞典. 三省堂, 2002.
- [33] 伊藤雅光. 計量言語学入門. 大修館, 2002.
- [34] 黒田航. 「純粹内観批判: 生成言語学の対抗馬であるだけでは認知言語学は言語の経験科学になれない. URL: <http://cslsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/critique-of-pure-introspection.pdf>, 2007.
- [35] 齋藤俊雄. 英語コーパス言語学: 基礎と実践. 研究社, 1998.
- [36] 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.
- [37] 水谷静夫. 数理言語学. 培風館, 1982.
- [38] 黒橋禎夫, 長尾眞. 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, pp. 115–118, 1997.