

徹底した用法基盤主義の下での文法獲得:

「極端に豊かな事例記憶」の仮説で描く新しい筋書き

黒田 航

独立行政法人 情報通信研究機構 知識創成コミュニケーション研究センター

1 はじめに

言語の知識が生得的かどうかは、今のところ事実に依存する経験的問題と言うより、依然として言語の知識の定義に依存する理論的問題(のまま)である。どんな知識が言語の知識として妥当な知識かどうかを検討せずに、言語の知識が生得的かどうかを論じるのは空虚であり、次の点には最大の注意を払っておくべきである: どんな形であれ「言語の知識はこれこれこういう知識である」と最初に決めてかかるのは、論点先取の危険を犯している。例えば、「統語構造が(二股枝分かれしかない)木構造だ」と決めてかかるのは論点先取の可能性がある。「統語構造=意味構造はイメージ図式だ」と決めてかかるのは論点先取の危険がある。これを慎重に避けない限り、「言語の知識が生得的か否か」という問題に実証的に答えを出すことはできるはずがない。木構造であれイメージ図式であれ、言語の知識の候補は、常に言語学者の思いこみでしかないかも知れない。言語の知識の実態は、言語学者が想定している以上に統計的で抽象的なもの、喩えるならば量子のように直観で実態を把握することが難しいものなのかも知れない。

以上の問題意識の下で、私がこの論考で目指すのは、[17]での主張を発展させ、用法基盤モデル(Usage-Based Model: UBM)[10]と互換性のある

言語の個体獲得のモデルの可能性を，R. Port [13] の極端に豊かな事例記憶 (extremely rich, exemplar memory: EREM) の仮説の下で極限まで推し進めたら，言語の個体獲得の問題がどんな姿になるかを素描することである．

2 「極端に豊かな事例記憶」の下での言語獲得

2.1 ヒトが言語を学んでいる時，何を学んでいるのか？

議論に先立って次のことは明記しておくべきである：

- (1) ヒト (特に幼児) が (新しい) 言語を学んでいる時，正確に何を，どう学んでいるのかは，実際にはわかっていない．

経験的にわかっているのは，ヒトの言語発達が喃語期 (0 語期) から一語期へ，一語期から二語期へ，二語期から多語期へという段階を経て，徐々に完成に至るということである (子供が多語期以降のいつ，大人と同質の文法 (チョムスキー派の言う「定常状態」) に到達するのは評価が一定ではなく，定説がないようだ) ．

言語の個体獲得/個体発達の一般的な説明では，幼児は最初は語を音素列と指示対象の組み合わせとして個別に獲得し，それから複数の語の結合法=統語論を学ぶ，というものである．だが，これは本当に正しい記述なのか？

問題は，子供がどうやって「語」の一定の配列が語の意味を超えた意味，超語彙的意味 (superlexical meaning) をもっていることを知るか，つまり，子供がどうやって統語論を「発見」するかが説明されていない点にある．ここに，統語論の「発見」の必然化の問題が生じる．これは T. Deacon [1] の脳と言語の共進化説の難点として指摘された点であり，個物の記号化能力の延長としての文法という定式化には不可避免的に発生するパラドックスである¹ ．

幼児がはじめは語の単位での指示しか理解できず，それより大きな単位での指示を理解できないと想定する限り，統語論は何らかの形で「発見」され

る必要がある。だが逆に、子供がはじめから語より大きな単位での指示が理解できているならば、統語論は発見される必要はない。つまり、次のことが真でなければならない:

- (2) 語がはじめから「状況レベルの意味」と「語に固有の統語論」の対として獲得される。
- (3) ただし、発達的な事実を辻褄を合わせるために「子供の心内に形成される語に固有の統語論の知識 K と観察可能な行動 B (\approx 産出される具体的形式) との間には乖離がある (と想定するが、これは特にアドホックな仮定ではない)。

これは逆説めいているが、真であることが期待される経験的主張である。子供が生得的に語より大きな単位での指示、状況指示を理解できていることを立証する十分な経験的証拠があるとは言えないが、これは問題のパラドックスをうまく避ける唯一の可能性であるように思う。これは後述の Wray [16] や Mithun [11] の提案する超語彙的単位の語に対する優先性の出発点となる点である。

子供がはじめから語より大きな単位での指示、すなわち状況指示を理解できているという可能性を肯定的に受け入れると、言語獲得の根本問題は次の形に変形される:

- (4) 語の意味と統語の獲得と、語より大きな単位 (例えば「文」) の意味と統語の獲得との関係はどうなっているか?

「語に固有の統語論」の一例が M. Tomasello [15] の動詞の島 (verb islands) であり、より一般的には Lexicalized Tree-Adjoining Grammar (LTAG) [6] が定義している (supertags という) 語彙化された表示の単位であろう。

以上のことが示唆するのは語彙的知識と文法的知識の連続性である。だが、その種の連続性を想定すればそれで文法の発達のパターンの説明に十分か? そうではない。表示の性質こそが問題の本質である。

2.2 言語の知識を可能にする記憶の性質

表示 (representation) を問題にするということは記憶 (memory) を問題にするということである。言語の記憶を問題にするということは、言語の経験の符号化 (encoding)、(記憶内容の) 保持 (storage)、(保持されている記憶の) 思い出し = 想起 (remembering) (≈ 記憶内の検索 (retrieval)) を問題にするということである。従来の言語学がこれらを誤った形で理解している可能性が高いことを以下で論じる。

2.2.1 言語 (という経験) の符号化と保持

言語学者にしっかり理解されているとは思えないが、言語の特定の構造 (例えば、統語構造、意味構造) のモデルを選ぶということは、特定の記憶のモデルを選ぶということでもある。記憶のモデルには様々な種類があり、どのモデルの選択するかは決して自明のことではない。この選択の微妙さが言語理論の構築に意味することは軽微ではない。このことは「言語の記憶が極端に豊かな事例記憶 (EREM) である」という R. Port [13] の最近の議論を通じて認識が始まったばかりの事柄だと言えるだろう。言語の知識と記憶の関係が彼の示唆する通りだとすれば、従来の言語構造の表示の理論は多かれ少なかれ見直しを迫られるのは確実である。

このような意識の下で私は以下で、EREM の想定の下で認知言語学で提案されている用法基盤モデル [10] を再解釈し、その結果を極端な用法基盤 (Extreme(ly) Usage-Based Model: EUBM) という名で素描する。

2.2.2 EUBM の特徴

EUBM の下では、ヒトは (少なくとも理論上は) 自分の聞いたり読んだりした表現 (これは必ずしも「文」とは言えない) を、全部、そっくりそのまま覚えている可能性が許されている。音素や形態素のような抽象的/スキーマ

的な表示は、そのような生の記憶にアクセスするためのインデックスということになる（アクセスを効果的にするために、統計的に支配的なパターンがインデックスに利用されることになる。これはスキーマ (schemas) という一種のメタ知識の必要性の一つの説明でありえる）。

なお、自分の聞いたり読んだりした表現を全部そっくりそのまま覚えているという想定が「直観に反する」という異議はもっともなものだが、「事実反した」ものだとは言えない。この点については、記憶内容の保持と想起は別である点に注意されたい。EREM の下で困難なのは想起であり、保持ではない。記憶が利用される時に常に想起の感覚が伴うとは限らない。記憶の大部分は（プライミングを例に出すまでもなく）意識下で働くものなので、「直観をもてない」ことは EUBM を拒絶するための強い理由にはならない（覚えと思い出しは互いに拮抗している別のシステムで、ヒトは誰でも驚くほど多くを覚えているが、そのほとんどを（適切な手がかりがないために）思い出せないでいると考えるべきだと私は思う）。

2.2.3 言語的記憶の想起の仕組み

言語の想起の単位は何か？ 語が音韻上の想起の単位であるというのは、ありそうなことである。だが、意味上の想起の単位は語か？ 一般にはそのように想定され、多くの言語理論がそのような想定の下で設計されているが、それは正しくない可能性が高い。実際、基本的な単位が語より大きな超語彙的単位であることは、コーパス言語学から得られた重要な知見の一つである「語の意味より、それを含む常套句の意味の方が優先される」という常套句の原則 (Idiom Principle) [14] から推測できる。

語が意味上の想起の基本単位ではないという見方は言語学内部では決して一般的ではないが、外部では徐々に市民権を得ている考えのように思う。例えば、Mithun [11] は Wray [16] を下地にしつつ、言語の起源が HmMMM (Holistic multi-modal manipulative musical) のようなものだったかも知れないと論じている。Mithun の議論には幾つか難点が指摘できるが、ばらばら

だった語を結びつける仕方を学んだところに言語の起源があるのではないという認識は正鵠を得たものである。全体が部分に優先する — 正確には常に全体が部分と同時に与えられるという性質が Wray と Mithun の言語の本質の議論の中核となるものである (認知文法流 [9] の言い方をすれば、これは全体と部分が同時に与えられる時には常に、全体がベースで部分がプロファイルになるような依存関係が存在するということである)。

更に言うと (実態を調べると容易にわかることだが) 言語の記憶の想起は並列、分散的である。これは言語の記憶の符号化が事例ベースで起こっているとすれば、不可避的な帰結である、

全体が部分と同時に与えられ、語が想起の単位ではないとすると、言語の意味論の基本原理の一つである構成性が成立する理由が明らかではなくなる。この理由から、言語の意味に関しては全体と部分が同時に与えられるという仮定は誤りであると批判されることがあるが、私にはこの批判は論点先取にしか思えない。というのは、意味の構築に関して、構成性原理が成立するというのは意味の理論の要請であり、事実とは言えないからだ。(仮に語義の組み合わせが構成的だとしても、語義の脱曖昧化の段階で非構成性が関与するならば、文意の決定全体の計算は非構成的なものとなる)。これは、言語学の教科書や論文に (繰り返し) 掲載されるような理想化され、単純化された文以外の、極く自然な表現の任意のものに対し、語用論と意味論の区別なしに、妥当な記述を与えようと試みたことのある者にとっては自明な事柄に属する。言語表現には一般に意味の構成性が成立しているように多くの言語学者が錯覚しているのは、彼らがそのような文しか扱わないからである。

私は「語より大きな、意味の喚起の基本単位」が存在すると想定し、問題を単純化するため、この単位が (近似的に) 「文」であると想定している。だが、これには理論的な困難も伴うことは白状しておきたい。

私が「文」と呼んでいるのは、実は「近似的に「文」としか呼びようのない、抽象的な単位」のことである。この単位の実態は (調査されていないため)、あまりよくわかっていない。談話分析での重要な成果の一つは、実際

の会話を仔細に観察してみると，そこには(生成)言語学者が「文」と呼ぶような単位は存在しない(か認定困難)ということであり，私が「語より大きな，意味の喚起の基本単位」を「文」と同一視する根拠は薄弱である．

ここで私が「近似的に「文」としか呼びようのない，抽象的な単位」という説明で意図しているのは，正確には語の(共)項構造が満足される最小の単位である²．この単位は談話の流れの中に連続して生起していなくてもよい³．ただし，これは言語学的にも認知科学的にも確立した説明とは言い難いので，問題の単位の実態が何であるかは，ここでは(疑似問題でなければ)未解決な問題であると言うに留める．

2.3 EUBM の下での新しい筋書き

以上のことから何が示唆されるか？ 以上の議論から浮上するのは，次のような言語の知識(≈文法)の個体獲得の新しい筋書きである：

- (5) 言語を獲得する子供は，豊かな記憶を下地にして，(おそらく数百万の桁の)夥しい数の具体的な形式 f とそれが使われる具体的な状況 s との対 (f, s) がレコードになっているような巨大なデータベースを発達させる(このようなことが可能なのは(言語の)記憶の基本的仕組みが並列，分散的だからである)．
- (6) 語 w (の意味 $m(w)$) はそのレコードの f (と対になっている $s = m(f)$) のインデックスにしかになっていない(従来 of 多くの言語理論が想定するように心内「辞書」にある「語」が幾つか組み合わせて文が作られる/生成されるのではなく，それが部分となるような全体 — 言語学者が近似的に「文」と呼んでいる単位 — が(EREMのおかげで)そのままそっくり記憶されているからである)．
- (7) 今までに経験したことのない新しい(近似的な意味での)文の認識は，それにもっともよく似た(つまり共通性が最大な)文からの意味，音

韻情報の転用=転化 (transfer) によって達成される (転用の際には複数の情報源からの多重継承 (=ブレンド [3]) が起こるのが常態である) .

- (8) ある話者 X にとって例えば Colorless green ideas sleep furiously が (文法的だろうと) 容認可能/理解可能でないのは, それが X が知っている, どんな意味の通る文 (例えば Colorful camouflaged insects fly amazingly) にも「似ていない」(と判断される) ためである .

以上の新しい筋書きが社会的脳の仮説 [2] から見えてくる言語の起源とどんな関係にあるかは今の時点ではっきりしないが, ヒトの社会性が音声形式と具体的な状況 s との対 (f, s) がレコードになっているような巨大なデータベースを発達させる淘汰圧になっている可能性は高いように思う . これは Mithun [11] の Hmmm の Holistic と manipulative の部分と特に整合する .

EUBM の導入は文法の生得性の基盤を必要最小限にする効果がある (実際, 普遍文法があることすら保証されない) . その帰結として, EUBM の下では言語の普遍的特徴 (language universals) の説明は難しくなる . 普遍文法で説明されて来た多くのことは, ヒトの記憶の仕組みと世界の客観的構造の知覚の仕組みによって説明される必要が責任が生じる .

2.4 まとめ

言語の知識の獲得が正確にどんなものなのかを明らかにするためには, それに先立って言語の記憶の認知科学的に妥当なモデル化の問題が解決されている必要がある . それなしには, どんな説明も論点先取である . 私はこの小論で, R. Port の音韻論上の議論を受けて「(極端に) 豊かな言語の記憶」の可能性を積極的に検討する必要性を論じ, その延長線上に EUBM を想定し, それが提供する文法の個体獲得の新しい筋書きを提示した . この論文は実証的なものというよりプログラムのであり, かつ福音主義的な響きもあるものであるが, これが今度の実証的な研究の足がかりになれば, 幸いである .

Notes

¹これ以外の論点の一つは再帰性 (recursion) である。ヒトの文法の重要な特徴として再帰を挙げる論者が少なくない [5] が、すでに学習によって再帰を再現するコネクショニスト模型があるので再帰が学習不能だという議論は成立しない。だが、それ以前に、私にはこの種の議論は根本的に破綻しているように思える。

第一に、ヒトがそもそも無限回の再帰を行なえるという経験的な証拠があるとは思われない。言語学で紹介されるのは、 x think (that) S のような動詞を使って数回の埋めこみの起こっている作成を示し、「(大前提) こういう文もヒトの言語で可能である。(小前提 1) 埋めこみの回数を更に増やせる。(小前提 2) その数に上限はない。だから、(帰結) ヒトの言語では無限回の再帰が可能だ」と言うことである(これを議論 A と呼ぶ)。だが、これは科学的な証拠ではなく、せいぜい逸話である。A の大前提は正しい。小前提 1 も条件つきでなら正しい。だが、小前提 2 は誤りである。何回かを正確に言うことはできないが、確実に上限はある。それには次の二つの理由がある。上限のない再帰の能力は生物学的、進化論的に「高きつく」。どんな種のどの個体にとっても使用から生じる利益を維持のための費用が上回るような能力を保持し続けることは不適応だと言える。ヒトの個体にとって実際に必要かつ有用なのは、十分な回数、例えばせいぜい 3 回程度までの再帰を模倣できる能力である。ただ、埋めこみの深さの増加は、ヒトの進化の過程で実際に起こった可能性はある。その深さに適応的価値があれば、より深い埋めこみができる個体が選択された可能性があるからだ。

次に(言語学の通説に反して) 基本的統語操作が代入 (substitution) だと考える経験的証拠はない。統語操作が代入だ想定すると再帰が必然化する。だが、生成文法が言う意味での「新しい形式 f^* 」の派生は、(A) f^* と適切な位置で部分文字列が一致する複数の既成の形式 $F = \{f_1, f_2, \dots, f_n\}$ を適当に選び、(B) F を統合 (=ブレンド [3]) すれば構成することができる。統合は多重継承の下での(素性の) 単一化にすぎないので、計算可能性は保証されている(実際、私の提唱した Pattern Matching Analysis (PMA) [7] は(計算機上で実装されてはいないが) 理論上は部分パターンの重ね合わせで、この意味での「統合」という操作を実装している)。

²共項構造の定義は [8] を参照されたい。

³これは私が PMA [7] で部分パターン(subpatterns) と呼んだものに実質的に等しい。

参考文献

- [1] T. W. Deacon. *The Symbolic Species: The Co-evolution of Language and the Brain*. W. W. Norton & Company, New York/London, 1997.
- [2] R. Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, Cambridge, MA, 1996.
- [3] G. R. Fauconnier and M. Turner. Conceptual integration networks. *Cognitive*

- Science*, 22:133–187, 1998.
- [4] A. D. Goldberg. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, IL, 1995.
- [5] M. D. Hauser, N. Chomsky, and W. Fitch. The faculty of language: What is it, and how did it evolve? *Science*, 298:1569–1579, 2002.
- [6] A. K. Joshi. Starting with complex primitives pays off: complicate locally, simplify globally. *Cognitive Science*, 28:637–668, 2004.
- [7] K. Kuroda. *Foundations of PATTERN MATCHING ANALYSIS: A New Method Proposed for the Cognitively Realistic Description of Natural Language Syntax*. PhD thesis, Kyoto University, Japan, 2000.
- [8] K. Kuroda, K. Nakamoto, and H. Isahara. Remarks on relational nouns and relational categories. In *Conference Handbook of the 23rd Annual Meeting of Japanese Cognitive Science Society*, pages 54–59, 2006.
- [9] R. W. Langacker. *Foundations of Cognitive Grammar, Vols. 1 and 2*. Stanford University Press, 1987, 1991.
- [10] R. W. Langacker. A usage-based model. In B. Rudzka-Östyn, editor, *Topics in Cognitive Linguistics*, pages 127–161. John Benjamins, Amsterdam/Philadelphia, 1988.
- [11] S. Mithen. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Weidenfeld & Nicolson, 2005.
- [12] J. M. Pine and E. V. M. Lieven. Slot-and-frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18:123–138, 1997.
- [13] R. Port. How are words stored in memory? beyond phones and phones. *New Ideas in Psychology*, 25(2):143–170, 2007. The original version, entitled “Words, symbols and rich memory”, available online at <http://www.cs.indiana.edu/~port/pap/Words.symbols.rich.memory.snglsp.Au%g8.pdf>.
- [14] J. M. Sinclair. *Corpus, Concordance, Collocation*. Oxford University Press, 1991.
- [15] M. Tomasello. *First Verbs: A Case Study of Early Grammatical Development*. Cambridge University Press, 1992.
- [16] A. Wray. *Formulaic Language and the Lexicon*. Cambridge University Press, Cambridge/New York, 2002.
- [17] 黒田 航. 認知言語学の言語習得へのアプローチ. 言語, 27(11):38–45, 1998. [原典版: <http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/Language-Acquisition.pdf>].

[18] 早瀬 尚子 and 堀田 優子. 認知文法の新展開. 研究社, 東京, 2005.