

「用法基盤」から「事例基盤」へ:

妥当な言語記憶のモデルを求めて

吉川 正人 (慶應義塾大学大学院)

1. はじめに

言語の用本基盤モデル(Usage-based Model of language: Langacker 1987, 2000, 2009; Kemmer & Barlow 2000)では、言語は具体的な言語経験から学習可能であり、言語の知識とは使用された具体的な言語表現を抽象化することによって得られたスキーマの膨大なネットワークであるとみなされる。このような考えは、言語知識を表現を生成する抽象的な規則や原理の体系(=文法)とみなしその獲得不可能説・生得説を唱える生成文法理論(e.g., Chomsky 1965, 1995)とは対照的なものである。

用法基盤モデル(以下 UBM)は、言語に特化した学習機構や処理システムを想定せず、一般的な学習メカニズムの範囲内で言語の習得を説明できる点で優れており、また実際の言語使用データに基づいて理論を構築できる点で堅実であると言える。

しかしながら、その詳細を子細に検討してみると、現行の UBM は二つの大きな問題を孕んでいることが明らかになる。本稿では、この二つの問題を明確化し、その解決策として、近年次第にその勢力を増しつつある、「事例基盤(exemplar-based)」の言語モデルを採用することの有効性を主張する。

二つの問題とは、以下のようなものである: 1) 学習メカニズムの根幹を成す「スキーマ」の抽出/学習プロセスの詳細が明らかでなく、特にその「始まりかた」を説明することが原理的に困難である; 2) 獲得した知識の「運用」面の議論が不十分であり、言語産出/理解の実情が正しく記述できていない可能性がある。事例基盤のモデルを採用すれば、このような問題には適切な説明が加えられる。

2. 言語の用法基盤モデルとその問題

ここでは、まず UBM の概要を簡単に紹介し、習得プロセスのモデルを提示した上で、上にみたその問題点を明らかにする。

2.1. 導入

Langacker (1987)は自身の考案した「認知文法(Cognitive Grammar)」という研究プログラムの中で UBM を提案した。

UBM の基本想定は以下のようなものである:

- (1) a. 我々の持つ言語知識は具体的な言語形式からなり、その中には所謂活用形(e.g., *toes, beads, walls*, etc.)と基本形(e.g., *toe, bead, wall* etc.)双方が含まれる;
- b. 名詞の複数形化のような一般規則は「スキーマ」として具体的な形式から抽出されると考える;
- c. 従って、我々の言語知識では抽象的なスキーマと具体的な言語表現は「共存」しており、
- d. 両者を異なった実体と考える理論は「“規則か一覧か”の誤謬 (rule/list fallacy)」を冒していることになる。(Langacker 1987: 27-29, 46)

これは言語知識の構造を素描したものと捉えられる。

2.2. 動的用法基盤モデル

Langacker (2000)は、言語知識の構造にとどまらず、それを獲得していくプロセスを捉えるモデルの必要性から、「動的用法基盤モデル (Dynamic Usage-based Model: DUBM)」を提案した。

DUBM の根幹をなす学習プロセスは、定着(entrenchment)と呼ばれる。定着とは、頻度学習によって繰り返し生起するパターンを確立した知識にする最も基本的なプロセスである。定着し「包装済み (pre-packaged)」となった事象は「ユニット (unit)」と呼ばれる (Langacker 2000: 3-4)。定着には、具体的で詳細な情報を持つ言語の使用場面から「選択的 (selective)」な抽象化を行うことが必要となる (Langacker 2009: 628)。

定着したユニットは「スキーマ (schema)」, 実際に生起する詳細で具体的な使用場面はその「事例 (instantiation)」と呼ばれる (Langacker 2009: 629)。スキーマとなる特定のパターンは具体的な使用場面の中に「内在する (immanent)」とみなされ、パターンの抽出は、当該パターンを内在する使用場面を繰り返し経験することで達成されると説明される (ibid.)。このプロセスは、図 1 のように図式化される:

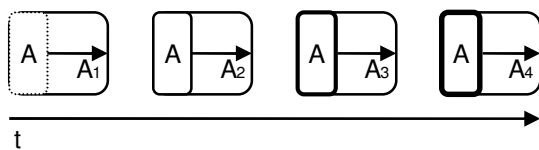


図 1. 定着の様子
(Langacker 2009: 630 の Figure. 2 を改変)

図 1 では、A がスキーマ、 A_i ($i=1\sim 4$) が A の事例を表わしている。事例の詳細は毎回異なるが、そこに内在する A は同一であり、その繰り返しの経験が A を定着させる。定着の度合いは枠線の太さによって表現されている。

Langacker (2009: 629) はユニット抽出のプロセスを以下のようにまとめている:

- (2) [T]he abstraction of units does not require any special mechanism; it is essentially automatic given the general Hebbian principle that occurring patterns of neural activation leave traces (in the form of strengthened synaptic connections) which facilitate their own recurrence.

2.3. スキーマ化はいかにして起こるか

しかしながら、このような説明には大きな問題がある。問題は、本当にパターンは具体事象に「内在的」なものであり、パターンの抽出はあるパターンを内在する事例を多数経験するだけでよいのか、ということである。

真の意味で「同じ」出来事は二度と起こらない。我々は何らかの方法で「(部分的)同一性」もしくは「類似性」を判断し異なる事象を「同じだ」と「看做している」わけである。本当に問うべきは、この「何らかの方法」である。しかし、その説明は UBM から決して提示されない。明らかに、UBM は「ヒトが類似性判断を行うことができる」という「事実」に基づいてモデルを構築している。

このことは非常に深刻なパラドクスを生む。図 1 に見た Langacker の考えるパターンの定着のプロセスには、奇妙なことに、初めからパターンの存在が前提とされている。しかし、パターンの存在は、それが具体的で直接経験可能でない以上、それを内在する具体的な事象を繰り返し経験することによって初めて「知る」ことができるものであるはずである。これでは、「パターンを抽出するには、そのパターンの存在を予め知っていなければならない」という事態が発生する。Yoshikawa (2009) は、これを「始まりのパラドクス (the beginning paradox)」と呼んだ。

誰もが実感しているように、実際の言語使用は、たとえ同一の語や言い回しであっても、その使われ方は一定ではない。残念ながら、音響的な観点からもこの同一性を規定するのは困難である (e.g., Port 2006, 2010; Johnson 2005,

2006)。UBM ではその同一性を保証する要素として「意味」を持ち出す (e.g., Langacker 2009: 628) が、仮にそうだとすると、言語を学習している幼児はどうやって意味の不変項に「気付く」のかということは不問にされている。これは音響的な不変項を発見することよりずっと難しいはずである。

2.4. スキーマはいかにして運用されるか

また UBM はスキーマを学習するプロセスに関しては詳細な検討がなされていると言えるが、その運用に関する議論はそれに比べると脆弱であるように思える。もし言語の知識が膨大なスキーマのネットワークであるならば、言語理解や産出はこのスキーマのネットワークを適切に運用することによって達成されることになる。

UBM では、理解/産出のプロセスは処理対象の事例と記憶している (= スキーマ) とのマッチングという形で達成されると看做される (e.g., Langacker 2009: 630-631)。これは言語理解/産出が常に「カテゴリー判断 (categorical judgment)」という形で達成されることを意味する。

これは正しい特徴付けだとは思われるが、問題は、このようなモデルではある言語処理における重要な側面を捉え切れない可能性が高いということである。それは、表現(の意味)の適格性判断における段階性である。話を言語理解に限定すれば、我々が何らかの表現を理解する時、それが何らかのスキーマの実現例であるということ認識するとともに、それがそのスキーマの実現例として典型的かどうかということも認識する。

UBM では、恐らくこの問題の答えをスキーマのネットワークの構造に求めるだろう。ネットワークの構造自体がカテゴリーの内部構造を表現しており、その中で成員の典型性も表現されると考えている可能性は高い。しかしながら、もしそうだとすれば、典型性/容認性の判断は離散的かつ非直感的・感覚的に行われるはずである。これは明らかに事実と反するであろう。

この問題は非常に重大である。単一の表現でも、使用される場面によって容認性は様々であるし、逆に言えば、柔軟に変わり得る場面・状況の変数をうまく扱える言語知識のモデルが不可欠であるということである。

3. 言語の事例基盤モデル

上記の問題を解決するには、「スキーマ」のような抽象度の高い宣言的な知識に依存するのではなく、もっと情報的に豊かで、可変的な振る舞いを許す実体に依存した言語知識のモデルを構築する必要がある。その有力な候補が、「事例基盤 (exemplar-based)」の言語モデル (e.g., Bod 2006; Gahl

& Yu 2006; Johnson 2005, 2006; Pierrehumbert 2001; Port 2007, 2010; Wedel 2006)である。¹

2006年に*The Linguistic Review*誌で特集されたことから分かるように、事例基盤モデルは近年注目を集めている言語モデルであると言える。以下で、事例基盤の言語モデル(EBM)がいかにしてUBMの難点を克服するのかを解説する。

3.1. EBMの基本想定

まずEBMの基本想定として、言語記憶は(抽象化無しの)エピソード的な非常に高次元(high-dimensional)の情報、即ち、「事例(exemplar)」の集積であると看做される。事例には無数の特徴量が素性として規定されており、実際の処理はこの素性を基に行われる。

このような記憶構造においては、抽象的な知識はそれ自体では存在していないことになる。基本的に抽象性は素性に基づく計算によって達成される。従って、スキーマのような(相対的に)低次元の構造は複数の素性の組織化の結果と看做される。また、もしある種の事例がカテゴリーのラベルと共に与えられたとすれば、カテゴリーのラベル自体も素性として付与されると考えられる。ただしそれは全ての事例に対して当てはまるわけではない。

3.2. EBMの想定する「スキーマ化」

スキーマの構造は記憶の実体としての事例の構造ではなく、そこに付与された(もしくはそれを構成する)素性の組織化の構造として表現されているため、EBMにおけるスキーマ化とは、複雑に組織化された素性の(ネットワーク)構造の出現と考えられよう。ここまでは特にUBMと相容れない想定ではなく、UBMの想定を詳細化、もしくは具体的なプロセスとして実装したものと捉えられるだろう。²

しかし、EBMはスキーマやパターンが事例のもつ素性の「共通項」の抽出によって「自動的」に得られるものだとはいえない。例えば音素のような抽象的な構造は音響的な特徴によって規定することが不可能であると主張されている(e.g., Port 2007, 2010)。これは(2)のような言明が単なる楽観視であり、そしてその時点でEBMはUBMと相容れない可

¹ Langacker (2009: 636)は事例基盤のモデルについて触れ、実際にはUBMの「代案(alternative)」ではなく、むしろUBMを実装するものであり、(等価ではないものの)非常に親和性が高い、ということ述べている。しかし、後に述べるように、事実は恐らくそうではない。

² これが脚注1で言及したLangacker (2009: 636)の真意であろう。

能性が高いことを意味する。

この含意として、抽象度の高いスキーマには心的な実在性はなく、我々が知識として持っているのはかなり具体的な、最低限の共通性を保証できるレベルのスキーマ/パターンであるという可能性が導かれる。Port (2007, 2010)の言うように、抽象度の高い低次元(low-dimensional)のパターン/スキーマは「認知的」に得られるものではなく、社会的な構築物である可能性がある。

またEBMでは言語事例にパターンが「内在している」とは考えない。パターンは蓄積された無数の事例(の素性)から「発見」されるものである。これはあらゆる事例がひとまず無節操に記憶されていると考えない限り辿りつけない想定である。

3.3. EBMによるカテゴリー判断

カテゴリー判断のモデルとしては、Hintzman (1986)の言う「スキーマ抽出(schema abstraction)」のプロセス、Johnson (2005, 2006)の「事例共鳴モデル(exemplar resonance model)」などが挙げられる。共通するのは、1) 事例に素性としてカテゴリーのラベルが含まれており、2) 入力となるカテゴリー判断の対象(Hintzman 1986の言う「標的(probe)」)と事例との類似度もしくは距離に基づく計算により、3) 事例に付与されたカテゴリーラベルが(重み付きで)得られる、というプロセスである。

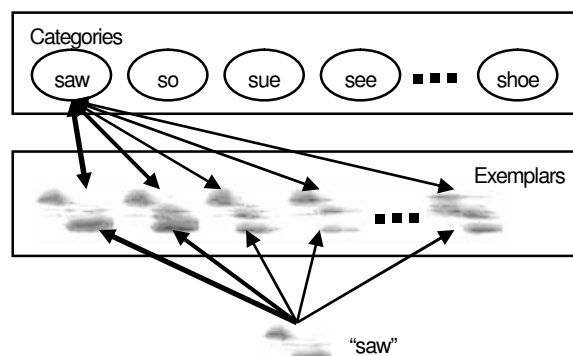


図 2. EBMにおけるカテゴリー判断の模式図 (Johnson 2006: 493のFig. 5を改変)

これはカテゴリーへの帰属度を自然に表す。また、入力となる対象に含まれる情報に応じて、選択されるカテゴリーラベルが変化するというのも自然に表現される。これは言語事例が発話の状況を丸ごとデータ化した高次元で豊かな情報を含んでいるというEBMの想定から導かれるものである。

尚、Pierrehumbert (2001)は同様のプロセスを実装したモデルによって産出のプロセスをシミュレートしている。彼女のシミュレーションは、発音面に関して徐々に安定したパ

パフォーマンスが得られるようになっていくプロセスを再現することに成功している。

3.4. 統語の獲得

このような議論を付き進めると、現行の UBM では統語(構造)の獲得が説明できないのではないかという可能性が浮上する。少なくとも、統語の知識を高度に抽象的な実体と考える以上、経験的にそれを獲得できるとするのは、上に見た議論から、非常に困難である。それは構文文法 (*Construction Grammar*: e.g., Goldberg 1995)を持ち出し、統語的な知識も統語パターンと意味の対である「構文 (constructions)」の集積と考えたとしても、この構文が高度に抽象的な表示 (e.g., Subject Verb Object₁ Object₂) であると仮定される以上、根本的な解決にはならない。

このことは、UBM による音韻・形態論の分析(e.g., Bybee 1995)が比較的盛んであるにも関わらず、統語論もしくは文形成論に関わる分析がほとんどないという事実に顕著に表れているように思える (ただし Bybee の研究には構成要素構造など句構造に関する議論は存在する: e.g., Bybee 2002; Bybee & Scheibman 1999)。

4. 結語

本稿では、現行の言語の用法基盤モデル (UBM) に対して、1) 言語知識の実体としてのスキーマを学習するプロセスが明確でなく、特に学習の始まり方を説明することが困難であり、2) スキーマの運用面に関しても議論が不十分で、言語理解/産出のプロセスを適切に説明できていない可能性がある、という難点を指摘し、それを克服するモデルとして、言語の事例基盤モデル (EBM) を紹介その利点について議論を行った。

EBM は計算言語学的な、また、自然言語処理的な側面を強く持つ。実際、EBM を採用したほとんどの研究でシミュレーションによる検証が行われている。その意味で、言語の記憶モデルとして EBM を採用することは、理論言語学と計算言語学および自然言語処理との融合を進め、新時代の言語研究への道を切り開く突破口となる可能性を秘めていると言える。

参考文献

- Bod, R. 2006. Exemplar-based syntax: How to get productivity from examples. *The linguistic review*, 23, 291-320.
- Bybee, J. 1995. Regular morphology and the lexicon. *Language and cognitive processes*, 10, 425-455.
- . 2002. Sequentiality as the basis of constituent structure. In Givón, T., & Malle, B. F. (eds.) *The evolution of language out of pre-language* (pp. 109-134). Amsterdam; Philadelphia: John Benjamins.
- Bybee, J., & Scheibman, J. 1999. The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, 37, 575-59
- Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, MA.: MIT Press.
- . 1995. *The minimalist program*. Cambridge, MA.: MIT Press.
- Gahl, S., & Yu, A. C. L. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The linguistic review*, 23, 213-216.
- Goldberg, A. E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago; London: University of Chicago Press.
- Hintzman, D.L. 1986. 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Johnson, K. 2005. Speaker Normalization in speech perception. In Pisoni, D.B. & Remez, R. (eds.) *The handbook of speech perception* (pp. 363-389). Oxford: Blackwell Publishers.
- . 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34, 485-499.
- Kemmer, S., & Barlow, M. 2000. Introduction: A usage-based conception of language. In Barlow, M., & Kemmer, S. (eds.) *Usage-based models of language* (pp. vii-xxii). Stanford: CSLI Publications.
- Langacker, R. 1987. *Foundations of cognitive grammar Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- . 2000. A dynamic usage-based model. In Barlow, M., & Kemmer, S. (eds.) (pp. 1-63).
- . 2009. A dynamic view of usage and language acquisition. *Cognitive Linguistics*, 23, 627-640.
- Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J., & Hopper, P. (eds.) *Frequency and the emergence of linguistic structure* (pp. 137-157). Amsterdam: John Benjamins.
- Pinker, S. 1991. Rules of language. *Science*, 253, 530-535.
- Port, R. 2007. How words are stored in memory: Beyond phones and phonemes. *New Ideas in Psychology*, 25, 143-170.
- . 2010. Rich memory and distributed phonology. *Language Sciences*, 32, 43-55.
- Tomasello, M. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA.: Harvard University Press.
- Wedel, A. B. 2006. Exemplar models, evolution and language change. *The linguistic review*, 23, 247-274.
- Yoshikawa, M., 2009. The "beginning paradox": An empirical and conceptual problem in the current Usage-based Model of language. *Colloquia*, 30, 57-74.