

乳幼児の語彙発達とダイナミックシステムアプローチ

阪上雅昭¹, 萩原広道²

¹ 京都大学 人間・環境学研究科

² 大阪大学 人間科学研究科

概要

乳幼児語彙の意味発達を検討するため、CDI データに VAE を適用し潜在空間を可視化した。英語と中国語の比較から、名詞と動詞等の分節化が異なる時期に生じることが明らかとなり、語の意味発達がアトラクター分岐として理解できる可能性を示した。

A Study on Infant Vocabulary Development based on the Dynamic Systems Approach

Masa-aki Sakagami¹, Hiromichi Hagihara²

¹ Graduate School of Human and Environmental Studies Kyoto University

² Graduate School of Human Sciences, The University of Osaka

Abstract

We applied a Variational Autoencoder to CDI vocabulary data to visualize lexical development. English and Chinese showed different timing in the emergence of noun-verb differentiation, suggesting that semantic development follows attractor-like bifurcations in latent space.

1 はじめに

発達初期の語彙は名詞が中心であることが広く報告されている。このため従来の研究は、名詞学習が早期に成立する理由や、動詞学習との困難度の差に焦点を当ててきた。しかし、語の意味がどのような過程を経て形成されるのかという問題は十分に解明されておらず、名詞が初期から静的な〈モノ〉カテゴリーを直接指示しているとは限らない可能性が指摘されている。

本研究が採用する胚性詞仮説 [1] は、語の意味発達を次の二点から捉える立場である。(a) 名詞の意味は初期には《靴を履く》のように〈モノ+行為〉が未分化な総体としての〈出来事〉に対応しており、〈モノ〉だけに分節化していない。(b) 発達のある時期を境に、語の意味は〈出来事〉から分節化(脱文脈化)し、名詞は〈行為〉に左右されない〈モノ〉のみを指示ようになる。この仮説に従えば、語の意味は〈出来事〉表象から〈モノ〉表象へと分化していく動的過程として理解される。

この意味分節化過程は、発達を非線形な変化として記述するダイナミックシステムアプローチ [2] と親和的である。同アプローチでは、発達は連続的な増加ではなく、状態空間におけるアトラクターの形成や

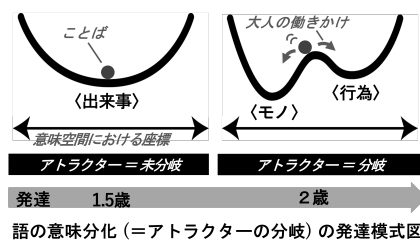


図1: アトラクターの分岐の模式図

分岐として特徴づけられる．胚性詞仮説をこの枠組みで捉えると，初期の〈出来事〉アトラクターから，名詞を中心とした〈モノ〉アトラクターと，動詞を中心とした〈行為〉アトラクターが分岐する過程として理解することができる（図 1）．

この仮説を検証するため，本研究では乳幼児の表出語彙を CDI（MacArthur Communicative Development Inventories）[3] により収集し，その高次元語彙データに対して変分オートエンコーダ（Variational Autoencoder：VAE）を適用した [4]．VAE は高次元データを低次元の潜在空間に写像する生成モデルであり，語彙の発達的变化が潜在空間においてどのような幾何構造を形成するかを明らかにすることが可能である．特に，潜在空間上で発達に伴いアトラクター様の構造が出現し，名詞と動詞の分節化に対応する分岐が観測されるかどうかを検討する．

2 VAE による乳幼児のことばの発達の可視化

図 2 は月齢 16 か月から 30 か月の乳幼児（米国）のべ 9093 名を対象に表出語彙数を調査した英語 CDI データを VAE[4] で解析した結果である．英語 CDI は，名詞が動物，おもちゃなど 11 カテゴリー，さらに動詞，形容詞，加えて代名詞や助詞などの閉じた語が 7 カテゴリーあり，全体で 22 カテゴリー 680 語のチェックリストである．こどもが表出している語を保護者がチェックリストに記入する形でデータが収集される．現在 42 言語についての調査結果が wordbank[3] というデータベースで公開されている．

本研究の VAE では，680 次元の CDI データを 3 次元の潜在空間に縮約し，その潜在表現から 680 次元データを再構成するように encoder と decoder からなるニューラルネットワークを教師なし学習で訓練した．encoder によって得られた被験者 9093 名の潜在空間での分布は，特徴的な「バナナ状」の構造を示した．さらに，英語 CDI の 680 語のうち，全く表出語彙がないケースを **birth point**，全語を表出したケースを **end point** とし，潜在空間上に配置した．birth point から end point に至る構造に沿って総語彙数が増加しており，被験者を総語彙数順に並べ 200 名ごとに平均位置を求めてつないだ緑の曲線は，潜在空間上の「**平均発達曲線**」を表す．このように潜在空間のバナナ構造は，乳幼児の語彙発達（表出語彙数の増加）の「かたち」と「うごき」を可視化している．

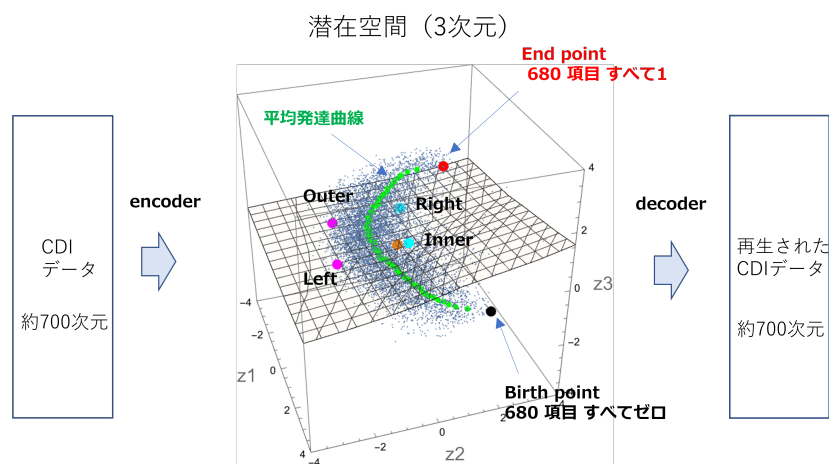


図 2: 英語 VAE と潜在空間

VAE は生成モデルであるため，潜在空間上の任意の点に対応する CDI データを decoder によって生成できるという特徴をもつ．その特徴を利用して，バナナ構造内の位置とことば発達の関係を調べた．図 2 のオレンジ点は潜在空間の原点を表しており，この点から見てバナナ構造の左側・右側を定義した．また，動径方向として原点から遠い方向を外側，近い方向を内側とした．図 2 の曲面は総語彙数割合が 0.4（総語彙数 $0.4 \times 680 = 272$ ）の等高面を示している．総語彙数が 272 語あたりの 200 名を抽出して共分散行列を計算し．左側，右側，外側，内側に 2σ ずらしたのが図中の Left, Right, Outer, Inner の 4 点で

ある．これら4点における680語の生成確率を decoder によって取得した結果，左側方向へのずれは名詞優位の語彙発達をもたらす，右側方向へのずれは動詞・形容詞・閉じた語といった動詞等優位の発達をもたらすことが示された．

外側・内側方向の違いには追加の説明が必要である．各語には vocabulary of acquisition (voa) が定義されており，これは平均発達曲線上でその語を半数の子どもが表出するようになる総語彙数割合を示す．総語彙数割合 0.4 の時点において，voa が 0.4 未満の語は半数以上の子どもが表出し，0.4 より大きい語は半数以下の子どもしか表出しない．潜在空間で外側方向にずれると，voa の小さい語（多くの子どもが一般的に早期に獲得する語）の表出確率がさらに高まることが分かった．これを本研究では「popularity が高い」と表現する．内側方向へのずれはその逆の傾向を示す．

他言語における発達構造がどのように現れるのかを検討するため，中国語 CDI データに対して同様の VAE 解析を行った．図3は，月齢16か月から30か月の乳幼児のべ1056名を対象とした中国語 CDI の潜在空間である．この潜在空間は英語とは大きく異なる特徴を示し，特に動径方向の厚みが非常に小さく，左右方向に大きく腕を広げた形状が得られた．このことは，中国語の語彙発達が英語と比べて，潜在空間の特定次元に強く集中的な構造をもつことを示唆している．

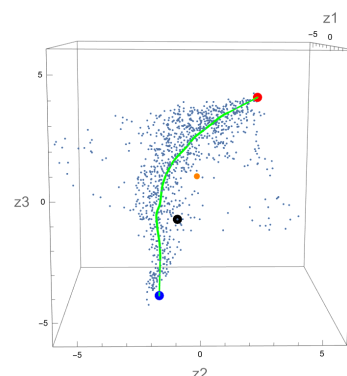


図3: 中国語 CDI の潜在空間

発達は英語の場合と同様，birth point から始まり，構造の下端（図中の青点付近）を通して end point に向かって進行する．緑の曲線は英語 CDI と同様に，総語彙数順に並べた被験者の平均位置を結んだものであり，（平均発達曲線）を表す．さらに，英語 CDI と同じ手法により，平均発達曲線からの偏差（左右・動径方向）が語彙発達に与える意味を調べたところ，中国語でも左側方向は名詞優位，右側方向は動詞等優位，外側・内側方向はそれぞれ popularity の高低に対応し，英語 CDI と同様の傾向が確認された．

これらの結果から，VAE によって得られる潜在空間は，CDI データに基づく乳幼児の語彙発達の全体構造を，相対論における時空図になぞらえることのできる「かたち」として提示していると言える．birth point から end point へ向かう軌跡が発達そのものを表し，潜在空間の位置が語彙選択の傾向を意味づけることも明らかとなった．加えて，英語と中国語では潜在空間の「かたち」が大きく異なっており，語彙発達の文化・言語的差異が潜在構造として反映されることは非常に興味深い．

3 VAE 特徴量による語の表出の発達の可視化

ここまでは，さまざまな発達段階にある子どもの CDI データを潜在空間に布置することで，語彙発達全体の「かたち」と「うごき」を可視化してきた．本章では，個体ではなく語そのものに焦点を当て，各語が潜在空間においてどのような発達の意味をもつのかを，VAE により定量化し可視化する．図4は語 go を例に，VAE 特徴量の算出方法を示したものである．

まず，go の voa (vocabulary of acquisition) は 0.125 であり，平均発達曲線上で総語彙数割合が 0.125 に到達した時点で，50 % の子どもが go を表出することを意味する．図中の曲面は総語彙数割合 0.125 の等高面である．この等高面上に描かれた曲線は go の表出確率の等高線であり，赤い矢印の方向に向かって単調に増加する．

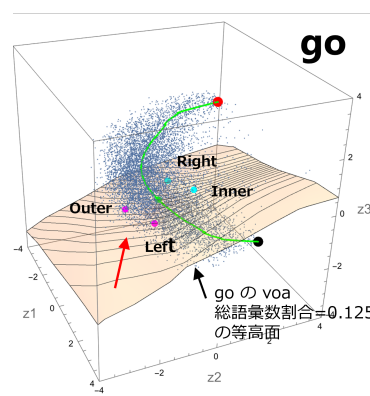


図4: go の特徴量の計算方法

Left, Right, Outer, Inner の4点は，第2章と同様に，平均発達曲線上の総語彙数割合 0.125 を保ったまま，潜在空間内で左側・右側・外側・内側にそれぞれ 2σ ずらした位置である．decoder によって求められた go の表出確率を比較すると，Left - Right = -0.407，Outer - Inner = -0.105 となる．これらの値は，それぞれ語がどの程度「名詞的方向／動詞的方向」に偏るか

(Left - Right), また「popularity の高低方向」に偏るか (Outer - Inner) を表す。したがって, (voa, Left - Right, Outer - Inner) = (0.125, -0.407, -0.105) が, 語 go の VAE 特徴量である。この結果は, go が比較的早期に獲得され (voa が小さい), 獲得時点では右側 (動詞等優位) 方向に強く, かつ内側方向にわずかに寄っていることを示す。

図5は英語 CDI の680語について同様に計算された3つのVAE特徴量の分布であり, クラスタリングによって3つのクラスターに分類されている。青色クラスターの語は CDI の名詞カテゴリーとほぼ一致し, これを名詞クラスターと呼ぶ。黄色および緑色のクラスターは, それぞれ動詞・形容詞カテゴリーおよび閉じた語カテゴリーに対応する。ここで強調しておきたいのは, クラスタリングには品詞情報を一切使用しておらず, VAE 特徴量という純粋に表出頻度に基づくデータ駆動的な指標のみを用いている点である。それにもかかわらず, 語の意味論のカテゴリーに対応する構造が自動的に抽出されていることはきわめて興味深い。

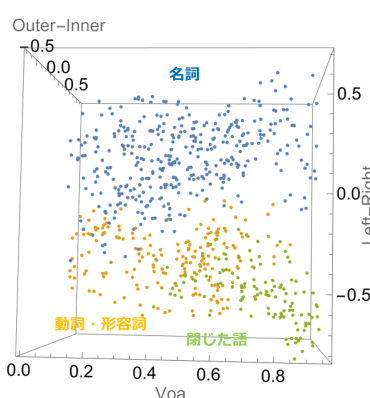


図5: 英語 VAE 特徴量

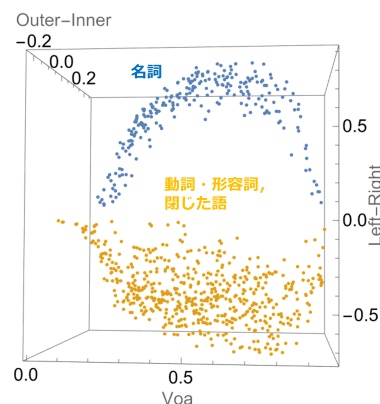


図6: 中国語 VAE 特徴量

図6は中国語 CDI の799語についてのVAE特徴量である。中国語の場合は, 語彙が名詞クラスターと動詞等クラスター (動詞・形容詞・閉じた語) に明瞭に二分される。さらに, 図6では voa が小さい発達初期の段階ですでにクラスターの分離が明確に現れている。一方, 図5の英語 CDI では発達初期にはクラスター境界が曖昧であり, 発達が進むにつれて両クラスターが徐々に分離していく様子が見られる。

4 まとめ

本研究では, 乳幼児語彙の発達過程をダイナミックシステムの観点から捉えるため, CDI データに対して VAE を適用し, 潜在空間における語彙発達の「かたち」と「うごき」を可視化した。その結果, 英語・中国語ともに, 潜在空間上で名詞優位方向と動詞等優位方向が明確に表れることが確認され, 語彙意味の分節化がアトラクターの分岐として理解できる可能性が示された。

特に, 中国語では発達初期の段階から名詞クラスターと動詞等クラスターが明瞭に二分されており, 名詞アトラクターと動詞等アトラクターが早期に形成されていることが示唆された。一方, 英語では初期段階においてクラスター境界が曖昧であり, 発達が進むにつれて徐々に分離が強まることが見られた。これは, 英語の語彙発達が胚性詞的な未分化状態から分節化へと移行する過渡期をより長く保持していることを示すものである。

参考文献

- [1] H. Hagihara and M. Sakagami, Cognition **226** (2022) 105177
- [2] E. Thelen and L.B. Smith, "A Dynamic Systems Approach to the Development of Cognition and Action", MIT Press, Cambridge, MA (1994)
- [3] M. Frank, et al. Wordbank: <https://wordbank.stanford.edu/>
- [4] D.P. Kingma and M. Welling, (2014). Auto-Encoding Variational Bayes.