

記号創発ロボティクスと マルチモーダルセマンティックインタラクション ～ 実世界認知・運動・言語を統べる知能構成への挑戦～

Symbol Emergence in Robotics and Multimodal Semantic Interaction

谷口忠大*¹ 岩橋 直人*² 新田 恒雄*³ 岡田 浩之*⁴ 長井 隆行*⁵
Tadahiro Taniguchi Naoto Iwahashi Tsuneo Nitta Hiroyuki Okada Takayuki Nagai

*¹立命館大学 情報理工学部

College of Information Science and Engineering, Ritsumeikan University

*²(独) 情報通信研究機構

National Institute of Information and Communications Technology

*³豊橋技術科学大学 大学院工学研究科

Graduate School of Engineering, Toyohashi University of Technology

*⁴(独) 玉川大学 脳科学研究所

Tamagawa University Brain Science Institute

*⁵電気通信大学大学院 電気通信学研究科

Faculty of Informatics and Engineering, The University of Electro-Communications

In this paper, a concept of symbol emergence in robotics and multimodal semantic interaction is described. Symbol emergence in robotics is a research field which treats real-world cognition, motion learning, vision learning, auditory learning, language acquisition and their emergence. We describe the reasons why we started this research field and our goals.

1. はじめに

人間知能は大きく記号的な処理に関わる上位知能と、身体的な動作に関わる下位知能に分けられると考えられてきた。しかし、それらは密接に繋がっている。我々は自己のセンサ・モータ系を通して構成される主観的な認識世界、つまりユクスキュルの述べる環世界 [1] の中で概念を獲得し、言語を獲得しコミュニケーションを行うに至るのである [2]。

記号処理を中心に構成された古典的な人工知能は現実世界での動作が困難であった。また、固定的な身体制御則や動作モジュールを設計しロボットに埋め込むことでは、なかなか人間の知能の振る舞いに到達しない。これらが示す事は、人間の知能は機能として設計するのではなく、環境との相互作用を通して、変容し育つべきものであるということであるが、その育つ機構の設計は容易ではない。しかし、近年、機械学習やロボティクスの技術的発展も相まって徐々にその可能性は広がりつつある。我々は記号的と言われる上位の知能が身体から出発し創発的に獲得されるボトムアップな知能構成へ挑戦したい。

言語や記号とは主体外部に真なる記号系が存在するものであるというのは、静的な環境下における近似に過ぎず、記号のなす表象機能や、サイン、また、その語間の関係としての統語も動的に変化するものである。音声認識も自然言語処理も知能ロボティクスもニューラルネットワークや進化計算などのソフトコンピューティング技術を経ながら、近年ではベイズ理論に基づく機械学習の恩恵を受け大いに発展している。しかし、その多くは教師ありの機械学習に基づきラベルや教師事例の与えられた限られたコーパス、データセットから学習するものである。そのラベルの準備や利用の枠組み構築に大きな人的労働コストがかかり、与えられたデータセットからの汎化の問題も含め、現実的には適用範囲が限られてしまうという問題があ

連絡先: 谷口忠大, 立命館大学情報理工学部, 滋賀県草津市野路東 1-1-1, 525-8577, 077-561-5839, taniguchi@ci.ritsumeikan.ac.jp

る。音声言語研究の現状を見ると、開発者は対話内容を直接記述しなければならないため、簡単なタスク下で自発的応答をしているかのように設計できるものの、労力負担が大きく、柔軟性に欠けるロボットしか制作できないというのが現状である。これは、対話に使用する背景知識の獲得と知識の運用方法に大きな課題が残されているためと考えられる。また、画像処理研究において、実世界環境の分節化や物体認識・学習などの実世界理解を、ロボットと人間のコミュニケーションタスクを相互作用の項を含めた最適化基準を用いて行う研究はまだほとんど見られない。

人間と意味的な関係性をつくるロボットは、その主観的な視座から自らのセンサ・モータ情報の分節化を経て記号的概念を獲得し、環境や他者との相互作用を通して意味づけを行わねばならない。さらに、獲得した記号を他者に対して用い、他者と相互信念を形成することで初めて、二者間の記号過程が生まれ、コミュニケーションは創発する。言語、運動、画像、音声、触覚などのマルチモーダルな情報を統一的に扱い、記号系の創発を通して意味の世界を析出する、身体性を伴った実世界認知・運動・言語を統べる知能構成への挑戦領域を、記号創発ロボティクスと呼称することとしたい。本稿では、そのコンセプトと研究課題について述べる。

2. 記号創発ロボティクス

2.1 記号過程と育てる設計論

記号創発ロボティクスは明示的に記号をとりあつかう。記号とは何かは記号創発ロボティクスにとって重要な問いである。記号論のパーズは記号過程 (semiosis = セミオシス) として記号を捉えた [3]。記号とは文字列の事ではなく、記号過程としての動的な認識のプロセスなのである。パーズは記号の三項関係によってこれを示した。記号過程はサイン (表象体)・対象・解釈項の三項関係で表される。ここで解釈項とは、サインと対象を結びつける心的な活動である。ここで重要なのはサ

インと対象の結びつきは恣意的なのであり、文化や文脈によって変わることにある。この点から、記号過程は常に自律的な環境認識活動の中で生まれるものなのである。例えば、夜に机の上に置かれたコップは誰かがねる前に牛乳を飲んだ事を示した記号かもしれないし、忘れずに薬を飲みなさいと言う妻からのメッセージかもしれない。このような記号過程の中から析出され洗練されたサインが言語なのであり、文字列そのものが記号ではないという理解は重要である。

記号過程を重視し、人間を要素として含むダイナミックなシステムの設計論の確立を目指す楳木は自律的な主体（ヒト、ロボット、細胞）が、他者を含む環境との相互作用を通して、意味の世界を創出して、伝達し、秩序が組織化される仕組みを解明することによって、外部からは境界条件のみを与え、あとはシステムを構成する主体群の固有のダイナミクスに沿わせて適応を実現することが必要であると論じる [4]。つまり、育てる設計論の確立が記号過程を内包したシステムにおいては重要なのである。

2.2 記号創発システム

筆者の一人である谷口はコミュニケーションするロボットを生み出すためには記号創発システムの構成論的な理解が不可欠であるとする [2]。記号は不変たりえず、文化や文脈に依存する。文化とはいわば、長期に染み渡った文脈であり、やはり固定されたものではない。ソシュールはこれを記号の恣意性と呼んだが、記号論研究の方法論の中ではその恣意性の人間知能の作動を基礎とした構成が十分に示されることは無かった。記号はその恣意性を前提とすると、認識主体の解釈に決して真なる答えが有るわけではなく、常に変容しうる。二者が対話するときも、お互いに暗黙的に前提を共有している際には、発話から語は減り、また、共有が不十分な場合には、過剰な記述がなされる。このような人間の認識活動、言語活動の中で揺らぐ記号を捉え、工学的にロボットに実装する為には、実世界の認知・運動と結びついた言語活動及びそれと繋がるマルチモーダルセマンティックインタラクションをボトムアップに捉え、総体として構成的理解を得る必要がある。

記号系は元来創発的なものである [5]。一言語の中でも、メディアが変われば、トピックが変われば用いられる語はかわり、その解釈はかわり、新語が生成される。記号創発を担う人間の知能を理解し、また、環境や人間との相互作用を通して自ら概念獲得し、人間と記号論的相互作用を行うロボットを生み出すためには、実世界の膨大なマルチモーダル情報から意味を析出し、対話を通じてその語の用法や効果を学習する過程を計算論的に理解する必要がある。機械学習の理論を基礎としながら、明示的な計算論として組み立てつつ、実世界で動作させることによって、筆者らは創発的に生まれる記号系のダイナミクスに接近することが出来ると考えている。このように記号創発システムにロボティクスの技術を用いつつ構成論的にアプローチし、またそれを通して多様な領域と相互作用することによって記号論的な人間理解を深め、また、真にコミュニケーション可能なロボットを生み出すのが記号創発ロボティクスである。我々はそれに挑戦するのである。

3. 研究の領域

記号創発ロボティクスに関わる研究を俯瞰し、当領域のこれからについて考察する。記号創発ロボティクスで扱うべき研究課題は以下に限らず、広い領域の一例と捉えるべきである。

3.1 マルチモーダル対話

我々は言語を文書としてのみ扱うわけではない。そこにはマルチモーダルな情報が関わってくる。岩橋らは相互信念モデルを組み込んだ言語獲得ロボットの開発を行いマルチモーダルな感覚情報を用いた学習と対話が可能であることを示してきた [6]。記号の持つ本質的な状況依存性を取り扱うためには、身体を持った主体同士の状況の共有と相互信念の形成が必要となる [7]。

また、ロボットは触覚や音声、画像などのマルチモーダル情報はお互いのモダリティの間に統計モデルを通じて意味づけを与え合うことが出来る。中村らはマルチモーダルな情報をロボット自身が言語と共に組織化させることによってモダリティ間のカテゴリ対応を得る事ができる事を示し、マルチモーダル情報から教師なし学習による言語の獲得の可能性を示している [8]。コンテキストに依存した意味処理や、言語・運動・画像の統合的な処理は本領域の重要な課題である。

3.2 概念獲得

人間が概念・言葉を獲得する際には様々なモダリティが介在する。これらのモダリティを通して構築される認識世界が主体にとっての環世界であり、その環世界に応じて世界は分節化される。谷口らは異なる身体、それ故に異なる環世界を持った主体が、異なる概念分化を起こす事についてシエマモデルを用いた構成論的研究を行い、そこに非線形性があることを指摘した [9]。Nakamura らは LDA に基づきマルチモーダル情報から概念化を行う手法を提案している [10]。人間は道具の概念を獲得する際にその持ち方や利用法といった行為を介した情報も積極的に得る。Nakamura らは道具の持ち方や機能、見た目の関係性を学習する概念モデルを構築しその有効性を示した [11]。

また、人間は新たな概念を獲得する際に、様々なバイアスに基づいて学習する事が知られており、またそれは発達によっても影響を受ける。例えば、あるオブジェクトを操作する動作を示した時、そのオブジェクトを中心に捉えるか、動作の軌跡自体を中心に捉えるかは年齢に依存する事が見いだされる [12]。この構成論的理解も重要な課題であろう。田口らは相互排他性バイアス、形状類似バイアスをエージェントに持たせる事によって概念獲得を加速させようという取り組みを行った [13]。対称性バイアスを始めとした多くのバイアスが知られており [14, 15]、これら人間が持つとされる言語獲得の特性がどのような計算論で説明しうるのかを構成的に示す事も記号創発ロボティクスの課題と言える。

3.3 言語獲得と発達

音声言語をボトムアップに獲得するには、音韻、単語に文法と、階層的な構造を如何に人間が教師なし学習を通じて形成しているかを計算論的に追う必要があるだろう。計算論的な古い常識では困難と思われるような、人間の言語獲得過程のプロセスもモデル化することにより、理解を深めていく事が出来る。

単語を為すチャンクの獲得については、指示対象の情報が一切無くともボトムアップに抽出しうる事が明らかになってきている。日下らは、音素列だけから単語や文法構造を MTRNN を用いる事によってボトムアップに抽出しうることを示した [16]。持橋らは二重分節構造を持った文章の完全な生成モデルを提案し、教師なしの形態素解析装置を開発し、教師あり学習に基づく手法に匹敵する精度を出すことを報告している [17]。また、音素列を抽出する音韻モデル自体を教師なしで学習することも研究されている [18]。

新規な語を獲得すること、人間にとっては無意識的に出来ることであるが、ロボットにとっては多くの困難がある。例えば、ある単語を聞いたときにそれが新語なのか既知して

いる物をさすのかを判定するという問題も重要である。岩橋は画像情報も用いる事によって、この問題を解決するモデルを提案している [19]。田口らは最小記述量基準を用いて連続的な発話から未知語を獲得しうることを示している [20]。このように、言葉を覚えていくプロセスの諸相をどのようにモデルを通して理解していくかも重要な問題である。

現在のところ、幼児のように音響データと他のマルチモーダル情報のみからボトムアップに音韻、単語、文法のモデルを形成し、新語を累積的に獲得するプロセスを一体的に表現するモデルは提案されていない。これを見出し、言語獲得と発達の過程の構造的な理解を得ることは重要である。

3.4 対話戦略

他者に話しかけるという行為も、決して乱数に支配されている訳ではない。プラグマティックに考えると、しばしば機能的な原理がその下に潜んでいる。人間が何らかの指示を受けた際に、それを聞き返すかどうかは、自らの認識に対する自信と、指示の結果実行可能な選択肢との相互作用から決定される。杉浦らはロボットに自身の認識結果の信頼度を自ら推定させ、聞き返しの調整を行う手法を示している [21]。また、そのような受け身の学習のみならず、積極的に発話を生成し、相互信念を深めようとする対話戦略を能動学習の理論をもとに形成している [22]。このような適応的な対話戦略についての研究はまだまだ少ないが、適切な設計論を構築していく必要がある。

3.5 動作や行為の学習

非分節な人間動作の観察を通して、ロボットが動作を獲得する研究もなされている。二重分節構造を持つことを前提とするならば、動素とその連なりとしての単位動作をボトムアップに抽出することができる [23]。動作がこのような構造を持てば、それをある種の言語と見なすことが出来る。高野らは動作の統計モデルと自然言語を翻訳の技術で結ぶことによって相互の変換を実現している [24]。一方で、尾形らは動作言語変換を RNN を用いて可能にしている [25]。これらのプロセスをボトムアップかつロバストに実現する事が出来れば、インタラクションを通じて様々な動作とそれを示す記述を獲得するロボットを作る事が出来ると考えられる。また、動作のみならず、様々な感覚情報は記号的モデリングが可能と考えられる。それらを扱う統一的な枠組みも望まれる。

ところで、観察を通じ学習された動作は調整されたり、汎化される必要がある。稲岳は得られた left-to-right 型の HMM で表現された単位動作のパラメータを外挿、内挿し調整する手法を提案している [26]。また、軌道として得られた「近づける」など一つの動詞で表現される動作も、何に近づけるのかで実際の動作としての表現は変わってくる。このように実空間での動作を言語のもとにどのように具体的に汎化させるかも、動作概念を獲得する上で重要である。杉浦らは参照点依存の動作生成を行う事で参照点を有する動作についての汎化問題を解いている [27]。また、学習された動作は合目的性をもって汎化され行為に変わる。この意味で記号性を有する動作・行為の意味学習も重要な課題であろう。

3.6 コミュニケーション・記号過程の創発

記号全体からみると、人間の言語は特殊事例とも言える。よりプリミティブなコミュニケーションを生む知能についての研究も重要である。記号の由来は何であるか。原初的にはオペラント条件付けにおける弁別刺激にあると考える事もできる。谷口らは STDP をモジュール型の強化学習と結合させた構成論的モデルを与え、記号過程の創発を創発の単純なモデルとして提案している [28]。日下らは RNN 学習器を用いて音と顔の向

きでインタラクションできる二体のロボットを向かい合わせる事によりコミュニケーションの創発を検討した [29]。高野らは二者間のインタラクションを統計モデルとして構成する HMM を結合することによって人間とロボットのインタラクションを実現している [30]。谷口らはインタラクションの中でロボットにインタラクションのルールを獲得させる手法として役割反転模倣手法を提案している [31]。実時間的な相互作用の中で、ロボットが自らの役割を発見したり、協調行動をいかに獲得するかも重要な問いである。また、いま・ここを超える言語の超越性をロボットにいかに持たせるかも大きな課題であろう。

4. 今後の展開

これらのように記号創発ロボティクスは身体から社会的な言語に至る多くの領域を総合的に対象にする。また、ここに挙げたもの以外でも多くの記号創発ロボティクスに関わる研究が発展的になされる事で、人間の本質的な知能についての理解が進展するものと考えられる。

記号創発ロボティクスの研究は、新たな挑戦であり、如何にしてボトムアップに言語的コミュニケーションを行う自律ロボットを生み出すかは開かれた問いである。その研究プラットフォームとして、筆者の一部はロボカップ@ホームにおいて実家庭環境でのタスク学習をロボットに行わせている。また、これらの研究の共通プラットフォームとしてエージェントシミュレータ Sigverse の開発されており、筆者らがのグループもその開発と研究に参画している [32]。これらの環境を含め多様な環境での研究を行う。

謝辞

本研究は国立情報学研究所共同研究助成「言語の超越性と記号創発に関する構成論的研究」「SIGVerse を用いたサービスロボットのための屋内外シミュレーション環境の構築とその応用」及び科学研究費補助金 学術創成「記号過程を内包した動的適応システムの設計論」19GS0208 の一部支援を受けた。

参考文献

- [1] ユクスキュル, クリサート. 生物から見た世界 (岩波文庫). 岩波書店, 2005.
- [2] 谷口忠大. コミュニケーションするロボットは創れるか - 記号創発システムへの構成論的アプローチ (叢書コムニス 13). エヌティティ出版, 2010.
- [3] C.S. パース, 内田種臣. 記号学 (パース著作集). 勁草書房, 1986.
- [4] 榎木哲夫. 記号過程を内包した動的適応システムの設計論 - つくる設計論から育てる設計論へ. システム/制御/情報, Vol. 54, No. 11, pp. 399-404, 2010.
- [5] 谷口忠大. 記号創発の意味がわからなくて困っています. (アイ・サイ問答教室). システム/制御/情報, Vol. 54, No. 5, pp. 214-215, 2010.
- [6] Naoto Iwahashi, Komei Sugiura, Ryo Taguchi, Takayuki Nagai, and Tadahiro Taniguchi. Robots that learn to communicate: A developmental approach to personally and physically situated human-robot conversations. In *Proceedings of the 2010 AAAI fall symposium on Dialog with Robots*, pp. 38-43, 2010.

- [7] 岩橋直人. ロボットによる言語獲得 – 言語処理の新しいパラダイムを目指して. 人工知能学会誌, Vol. 18, No. 1, pp. 49–58, 2003.
- [8] Tomoaki Nakamura, Takaya Araki, Takayuki Nagai, and Naoto Iwahashi. Grounding of word meanings in lda-based multimodal concepts. *Advanced Robotics*, to appear.
- [9] 谷口忠大, 榎木哲夫. 身体と環境の相互作用を通じた記号創発: 表象生成の身体依存性についての構成論. システム・制御・情報学会誌, Vol. 18, No. 12, pp. 440–449, 2005.
- [10] 中村友昭, 長井隆行, 岩橋直人. ロボットによる物体のマルチモーダルカテゴリーゼーション. 電子情報通信学会論文誌 D, Vol. J92-D, No. 10, pp. 2507–2518, 2008.
- [11] 中村友昭, 長井隆行. 機能と視覚的情報の関係性に基づく物体の概念モデル. 情報処理学会論文誌, Vol. 51, No. 8, pp. 1413–1427, 2010.
- [12] Etsuko Haryu, Mutsumi Imai, and Hiroyuki Okada. Object similarity bootstraps young children to action-based verb extension. *Child Development*, 2011.
- [13] 田口亮, 木村優志, 小玉智志, 篠原修二, 入部百合絵, 桂田浩一, 新田恒雄. 幼児の学習バイアスを利用したエージェントによる語意学習の効率化. 人工知能学会論文誌, Vol. 22, No. 4, pp. 444–453, 2007.
- [14] Mutsumi Imai, Sotaro Kita, Miho Nagumo, and Hiroyuki Okada. Sound symbolism facilitates early verb learning. *Cognition*, Vol. 109, pp. 54–65, 2008.
- [15] 今井むつみ, 岡田浩之. 特集「対称性」へのコメントリー: 言語の成立にとって対称性はたまごかにわとりか. *Cognitive Studies*, Vol. 1, No. 1, pp. 1–12, 1994.
- [16] Wataru HINOSHITA, Horiaki ARIE, Jun TANInd Hiroshi G. OKUNO, and Tetsuya OGATA. Emergence of hierarchical structure mirroring linguistic composition in a recurrent neural network. *Neural Networks*, Vol. 24, No. 4, pp. 311–320, 2011.
- [17] 持橋, 山田, 武士, 上田修功. ベイズ階層言語モデルによる教師なし形態素解析 (言語モデル・ウェブ解析). 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 2009, No. 36, p. 49, 2009.
- [18] N. Iwahashi. Language acquisition through a human-robot interface by combining speech, visual, and behavioral information. *Information Sciences*, Vol. 156, No. 1-2, pp. 109–121, 2003.
- [19] Naoto Iwahashi. Interactive learning of spoken words and their meanings through an audio-visual interface. *Transaction IEICE*, Vol. E91-D, No. 2, pp. 312–321, 2008.
- [20] 田口亮, 岩橋直人, 船越孝太郎, 中野幹夫, 能勢隆, 新田恒雄. 統計的モデル選択に基づいた連続音声からの語彙学習. 人工知能学会論文誌, Vol. 25, No. 4, pp. 5491–5501, 2010.
- [21] 杉浦孔明, 岩橋直人, 柏岡秀紀, 中村哲. 言語獲得ロボットによる発話理解確率の推定に基づく物体操作対話. 日本ロボット学会誌, Vol. 28, No. 8, pp. 978–988, 2010.
- [22] Komei Sugiura, Naoto Iwahashi, Hisashi Kawai, and Satoshi Nakamura. Active learning for generating motion and utterances in object manipulation dialogue tasks. In *Proceedings of the 2010 AAAI Fall Symposium on Dialog with Robots*, pp. 115–120, 2010.
- [23] 谷口忠大, 岩橋直人. 複数予測モデル遷移の n-gram 統計に基づく非分節運動系列からの模倣学習手法. 知能と情報 (日本知能情報ファジィ学会論文誌), Vol. 21, No. 6, pp. 1143–1154, 2009.
- [24] 高野渉, 山根克, 中村仁彦. 運動記号と運動ラベルの連想モデルに基づく運動データの検索・生成計算. 日本ロボット学会誌, Vol. 28, No. 6, pp. 723–734, 2010.
- [25] Tetsuya OGATA, Masamitsu MURASE, Jun TANI, Kazunori KOMATANI, and Hiroshi G. OKUNO. Two-way translation of compound sentences and arm motions by recurrent neural networks. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2007)*, pp. 1853–1863, 2007.
- [26] 稲岳哲也, 柴田智広. 動作パターンとシンボルを相互変換する原始シンボル空間における動作パターンの内挿・外挿. 日本ロボット学会誌, Vol. 28, No. 4, pp. 512–521, 2010.
- [27] Komei Sugiura and Naoto Iwahashi. Learning object-manipulation verbs for human-robot communication. In *Proceedings of Workshop on Multimodal Interfaces in Semantic Interaction*, pp. 32–38, 2007.
- [28] T. Taniguchi and T. Sawaragi. Incremental acquisition of behaviors and signs based on reinforcement learning schema model and stdp. *Advanced Robotics*, Vol. 21, No. 10, pp. 1177–1199, 2007.
- [29] 日下航, 尾形哲也, 小嶋秀樹, 高橋徹, 奥乃博. Rnn を備えた 2 体のロボット間における身体性に基づいた動的コミュニケーションの創発. 日本ロボット学会誌, ロボティク・サイエンス論文特集号, Vol. 28, No. 4, pp. 532–543, 2010.
- [30] 高野渉, 山根克, 杉原知道, 山本江, 中村仁彦. 身体的記号化モデルに基づく人間とヒューマノイドロボットのコミュニケーション理論. 日本ロボット学会誌, Vol. 28, No. 6, pp. 735–745, 2010.
- [31] 谷口忠大, 岩橋直人, 中西弘門, 西川郁子. ヒューマン・ロボットインタラクションを通じた役割反転模倣に基づく実時間応答戦略獲得. 第 23 回人工知能学会全国大会, 2009.
- [32] T. Inamura, T. Shibata, H. Sena, T. Hashimoto, N. Kawai, T. Miyashita, Y. Sakurai, M. Shimizu, M. Otake, K. Hosoda, S. Umeda, K. Inui, and Y. Yoshikawa. Simulator platform that enables social interaction simulation sigverse: Sociointelligence simulator. In *2010 IEEE/SICE International Symposium on System Integration*, pp. 212–217, 2010.