2.1.2 言語・知識系のAI技術

（1）研究開発領域の定義

知能を知覚・運動系と言語・知識系という2面で捉え、ここでは後者を俯瞰する（前者については前節2.1.1で俯瞰した）。研究開発領域としては、自然言語の解析・変換・生成等を行う自然言語処理（Natural Language Processing）、知識の抽出・構造化・活用を行う知識処理（Knowledge Processing）等が中心的に取り組まれてきた。

知覚系は実世界からの入力、運動系は実世界への出力として、知能の実世界接点の役割を担う。状況を知り、判断し、行動するという一連のプロセスは、知能において、知覚系と言語・知識系と運動系の連携によって熟考的に実行されることもあれば（ここでは熟考的ループと呼ぶ）、知覚系と運動系の間で即応的に実行されることもある（ここでは即応的ループと呼ぶ）。近年、機械学習（Machine Learning）、特に深層学習（Deep Learning）が発展し、まずは知覚系（パターン認識）での活用が進み、次第に言語・知識系（自然言語処理、知識処理）や運動系（動作生成）にも広く活用されるようになった。知覚・運動系と言語・知識系の処理方式の共通性が高まり、それらを統一的に扱う枠組みも研究されるようになってきた。そこで、本節では、自然言語処理・知識処理そのものの研究開発の動向に加えて、知覚・運動系と言語・知識系を統合して熟考的ループを構成するための研究開発の動向についても取り上げる。

（2）キーワード

自然言語処理、知識処理、テキスト処理、機械学習、深層学習、分散表現、アテンション、トランスフォーマー、意味理解、文書読解、質問応答、機械翻訳、文章生成、知識獲得

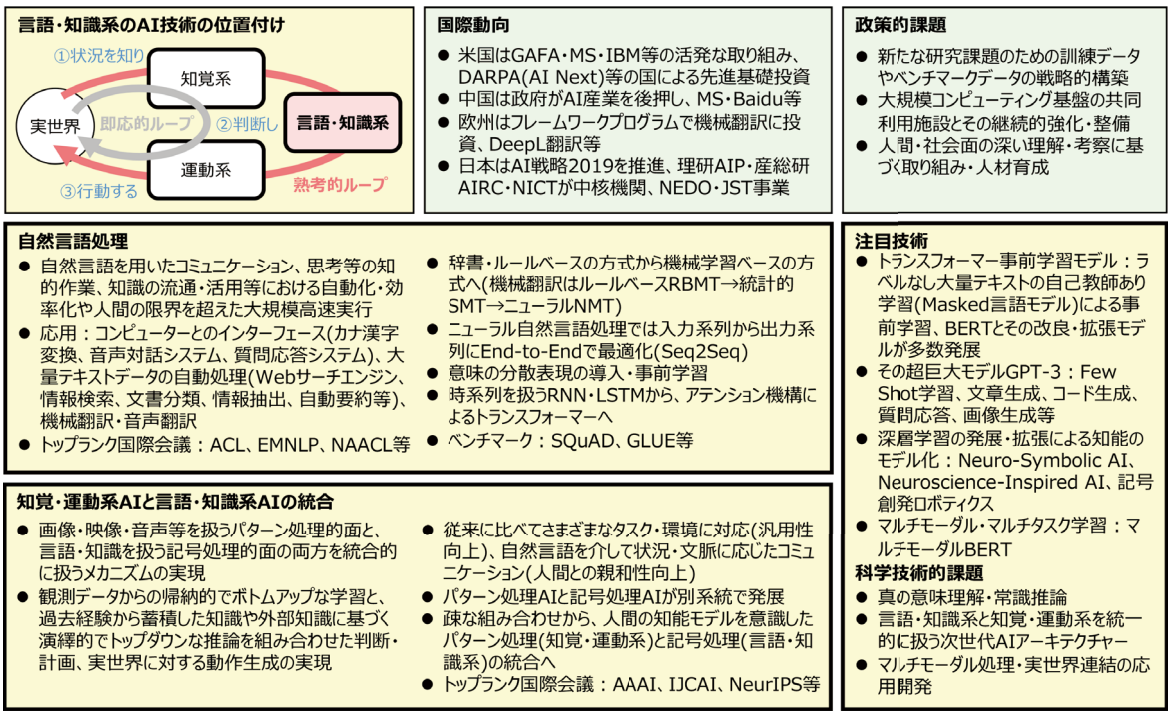

図2-1-4　　領域俯瞰：言語・知識系のAI技術

（3）研究開発領域の概要

［本領域の意義］

自然言語は人間が日常の意思疎通のために用いる自然発生的な記号体系である[1]。人間にとって自然言語は、概念を表現する記号体系として、日常の意思疎通（コミュニケーション）だけでなく、思考の過程やその結果である知識の表現・保存にも用いられる。コンピューターによる自然言語処理と知識処理は、このような人間のコミュニケーション、思考等の知的作業、知識の流通・活用等を含むさまざまな場面に適用され得る。そして、その自動化・効率化や、人間の限界を超えた大規模高速実行を可能にする。その代表的な場面・システムのいくつかを以下に挙げる。

まず、コンピューターとのインターフェースに使われる自然言語処理として、カナ漢字変換入力システム、音声対話システム、質問応答システム等が挙げられる。最近はスマートスピーカー（Amazon Echo、Google Home等）が家庭で使われ始めているが、自然言語で操作・指示できると、特別のコマンド入力・操作方法をあれこれ覚える必要がない。コンタクトセンターでの問い合わせ受付では、簡単な質問への対応の自動化によって、問い合わせ対応のスループット向上や質の安定が得られる。

また、大量のテキストデータの処理をコンピューターで行うことで、人間の負荷を軽減しようという自然言語処理システムがある。 Webサーチエンジンが代表例だが、膨大な情報の中から条件に合う情報を高速に見つけたり、整理したりするための情報検索・文書分類、その概要把握を助ける情報抽出・自動要約等に、自然言語処理が活用され、例えば、科学技術研究の加速につながっている。また、大量のWebテキストから、概念の間の関係をリンクで表現した大規模なナレッジグラフ（知識グラフ、Knowledge Graph）を構築し、検索語の拡張、検索結果の品質向上、対話システムの話題拡大等に利用することも行われている。

さらに、機械翻訳・音声翻訳（自動通訳）も自然言語処理の代表的な応用である。母国語から他国語、あるいは、逆に他国語から母国語への翻訳・通訳は、他国語を話す人々とのコミュニケーションを支援するとともに、インターネット等を介して世界に流通している膨大な量の他国語で書かれた情報を調査・分析する労力を大幅に軽減してくれる。

以上、自然言語処理・知識処理の意義や応用について述べたが、次に、本節で取り上げるもう一つのトピックである知覚・運動系と言語・知識系の統合（熟考的ループ）について、その意義や応用について述べる。この統合によって、人工知能（AI：Artificial Intelligence）やロボットを実現する上で、より総合的な知能の性質がカバーされる。すなわち、画像・映像・音声等を扱うパターン処理的な側面と言語・知識を扱う記号処理的な側面の両方を統合的に扱うメカニズムが実現される。また、知覚系を通して直接的に得られる外界の観測データからの帰納的でボトムアップな学習と、過去の経験を通して蓄積された知識や社会・他者と共有された外部知識に基づく演繹的でトップダウンな推論の、両方を組み合わせた判断や計画のメカニズムと、その結果を実世界に対する一連の動作として生成・実行するメカニズムが実現される。このようなメカニズムを備えたAI・ロボットは、従来に比べて、さまざまなタスクや環境に対応可能になり（汎用性の向上）、自然言語を介して、実世界の状況・文脈に応じたコミュニケーションが可能になる。このことは、人間との親和性を向上させ、人間とAI・ロボットが協働する中で共に知識を創成し、共に成長する社会の実現につなが

1 自然言語に対して、プログラミング言語やマークアップ言語等、人工的に定義された言語がある。これらの人工的に定義された言語は解釈が一意に定まるように設計されているが、自然言語は文・句・単語等の意味や構造の解釈に曖昧性が生じ得る点、記号接地（記号と実世界における意味をどのようにして結びつけるか）や意図理解のように記号だけに閉じない問題が関わる点等が、その処理を難しいものにしている。

ると期待される。

［研究開発の動向］

❶ 自然言語の解析技術の発展（～2017年頃）

自然言語処理技術において共通的に必要とされる基礎技術はコンピューターによる自然言語解析であり、形態素解析（単語分割や品詞認定）、構文・係り受け解析、文脈・意味解析（語義の曖昧性解消や照応解析を含む）というステップで、より深い解析への取り組みが進められた。そのアプローチは、黎明期の1950年代から1990年代頃まで、人間が記述した辞書・文法を用いるルールベース方式が主流だった。しかし、大量のテキストデータが利用可能になったことや、機械学習技術が大幅に進化したことから、徐々に統計的な方式、機械学習を用いた方式に主流が移った。適用される機械学習技術は、ナイーブベイズに始まり、2010年頃にはSVM（Support Vector Machine）が主流となったが、2014年頃からはニューラルネットワークによる機械学習、特に深層学習が盛んに適用されるようになった[1), 2), 3)]。

このような技術発展は特に機械翻訳への取り組みによって牽引されてきた。機械翻訳方式は当初のルールベース機械翻訳（Rule-Based Machine Translation：RBMT）から、1990年代に大規模な対訳コーパス（元言語のテキストとターゲット言語のテキストを対にしたもの）と機械学習技術を用いた統計的機械翻訳（Statistical Machine Translation：SMT）に主流が移った。 SMTの精度改善に頭打ちが見えてきた2010年代に、ニューラル機械翻訳（Neural Machine Translation：NMT）[2), 4)] が考案され、顕著な精度改善がもたらされた。 SMTからNMTへの移行は、SMTで用いていた統計処理・機械学習のパートを単純に深層学習に置き換えたものではなく、機械翻訳のパラダイムを大きく転換させたものである。SMTでは、機械翻訳のプロセスを多段階に分け、各段階の処理モデルを統計的にチューニングして組み合わせていたのに対して、NMTでは、入力原文から翻訳結果の出力までを1つのニューラルネットワーク構造（Seq2Seqモデル[5)]）として扱い、End-to-Endの最適化を行う。

その際、自然言語の単語系列をニューラルネットワークで扱うため、意味の分散表現[2]が用いられる。これは単語・句・文・段落等の意味を固定長ベクトル（実際には数百次元程度）で表現したものである[6), 7), 8), 9)]。大量テキストにおける文脈類似性に基づき、ニューラルネットワークを用いて分散表現を高速に計算するWord2Vecが2013年に公開され、自然言語処理の基本的な手法として広く使われるようになった。それまで使われていたBag-of-Words形式（N次元のうちの1要素だけの値が「1」というone-hotベクトル）と異なり、分散表現はベクトル計算によって単語や文の意味の合成・分解や類似度計算が可能である。例えば、分散表現を用いると、「king」-「man」+「woman」=「queen」のようなベクトル計算が可能である。従来の記号処理は厳密な論理演算をベースとした硬いものだったが、分散表現を用いることで曖昧な条件を許した柔らかい演算が可能になった。

2 分散表現（Distributed Representation）は、ニューラルネットワーク研究の分野では局所表現（Local Representation）に対する概念として考えられた。一方、自然言語処理研究においても、単語の意味を扱う方法論として分布仮説（Distributional Hypothesis）があり、これら両面が融合したものと考えられている[9)]。自然言語に限らず、なんらかの離散的な対象物の表現方法として、局所表現や分散表現を用いることができる。局所表現はone-hotベクトルのように1つないしは少数の要素で特徴を表現するのに対して、分散表現は多数の要素に特徴を分散させて表現する。また、埋め込み（Embeddings）という言い方も用いられる。たとえばDistributed Representation of WordsとWord Embeddingsは同義である。

この分散表現とSeq2Seqモデルを用いてEnd-to-Endで最適化するアプローチは、
入力系列（End）→［エンコーダー］→ 分散表現 →［デコーダー］→ 出力系列（End）
という流れになる。このような系列変換は、機械翻訳だけでなく、質問応答、対話、情報要約、画像・映像に対する説明文生成等、自然言語処理のさまざまな応用に使われるようになった[10)]。

「2.1.1 知覚・運動系のAI技術」で述べたように、深層学習はまず画像認識・音声認識の分野に適用され、衝撃的な性能向上がもたらされた。それに対して自然言語処理の分野では、当初、画像認識・音声認識分野のような華々しい性能向上が一気に得られたわけではなかった。しかし、この数年で新たな技術発展が生まれ、自然言語処理においても著しい進展がもたらされた。その内容は❸で後述する。

❷ 大規模テキスト活用・知識活用の発展

テキスト検索は、コンピューターの処理性能が乏しかった時代、事前に人手で各テキストに付与したキーワードを索引に用いるしかなかったが、1990年代以降、コンピューターの性能向上、並列処理技術の発展、ストレージの大容量化等が進み、フルテキストサーチ（全文検索）方式に主流が移った。急激に大規模化したWebサーチエンジンが、その代表であるが、クエリーのキーワードとWebのフルテキストの単純なマッチングでは高い検索精度が得られない。 Webの被リンク関係やアンカー文字列（リンク元テキスト）を考慮した検索結果のランキング法や、ユーザーの嗜好や目的に応じた適合ページの選別法等、さまざまな観点からWeb検索の精度を高める技術が開発された。また、大規模なWebを解析・検索するため、大規模自然言語テキストを解析・検索するための分散・並列処理、圧縮・索引処理等の技術が急速に発展した[11)]。Webサーチエンジンは幅広い一般ユーザー向けのアプリケーションとして発展したが、インターネット上の多様な情報や企業内の大量文書から評判・意見、注目事象、傾向変化等を抽出し、企業経営、マーケティング、リンク管理等に活用するテキストマイニング・Webマイニングと呼ばれる技術・アプリケーションも開発が進んだ[11), 12)]。また、Google は Web上の情報から大規模なナレッジグラフを構築し、検索結果とともに表示することを行っている。

さらにその発展として、大量のテキスト情報を知識源として用いる質問応答システムがある。代表的なシステムとしてIBMのWatson[13)] があげられる。 Watsonは、大量テキスト情報を知識源として自然言語で書かれた質問に回答する技術を中核とし、2011年に米国の人気クイズ番組「Jeopardy!」で人間のクイズ王に勝利するというグランドチャレンジに成功した。国内では、情報通信研究機構（NICT）が開発・公開したWISDOM Xは大規模なWeb情報をもとに、自然言語による「なに？（いつ／どこ／だれ）」「なぜ？」「どうなる？」「それなに？」という4タイプの質問に回答する。また、WISDOM Xの技術を応用した対災害SNS情報分析システムDISAANA、災害状況要約システムD-SUMMも開発された。災害時には、多様な災害関連テキスト（SNS、ニュース、報告書等）がバースト的に流通するが、厳しい時間制約・人資源制約の下、それらを精査・統合して適切なアクションを決定することは非常に難しい。 DISAANAは2016年の熊本地震の際に政府で活用され、自然言語処理技術が災害時の迅速な意思決定にも活用可能であることを示した。

❸ ニューラルネット自然言語処理の最新動向（2017年～2020年）

深層学習による画像認識には畳み込みニューラルネットワーク（Convolutional Neural Network：CNN）が主に用いられたが、自然言語処理には、時系列情報を扱うのに適した回帰型ニューラルネットワーク（Recurrent Neural Network：RNN）やLSTM（Long Short-Term Memory）ネットワークが用

いられた[9)]。これを用いた系列変換（Seq2Seq）の際に、ニューラルネットワーク中のどこ部分（特定の単語等）に注目するかを決定するアテンション（Attention）機構が考案され、出力系列生成の品質向上につながった。アテンションのアイデアは最初、2014年に機械翻訳に導入されたが[14)]、その後、自然言語処理全般で（さらには画像処理にも）用いられるようになった[4), 15)]。

このアテンション機構を最大限に生かした新しい深層学習モデルとして、2017年にトランスフォーマー（Transformer）がGoogleから発表された[16)]。トランスフォーマーは、RNNやCNNを使わずに、アテンション機構[3]のみで構成した深層学習モデルである。RNNやCNNより計算量が抑えられ、並列処理もしやすく、複数の言語現象を効率良く扱えて、文章中の長距離の依存関係も考慮しやすいといった特長を持ち、機械翻訳等のベンチマークでも従来を上回る性能が示された。このことから、2018年以降、新たに提案されるモデルはトランスフォーマー一色となり、次に説明する自己教師あり学習による事前学習の手法と合わせて、自然言語処理のさまざまなベンチマークで最高スコアが更新されている[17)]。

ニューラルネット自然言語処理で高い精度を達成するには、大量の訓練データが必要だが、さまざまなタスクの各々について大量の訓練データを用意することは容易なことではない。そこで、まず、さまざまなタスクに共通的な汎用性の高いモデルを、大量のラベルなしデータで事前学習（Pre-Training）しておき、それをベースに個別のタスク毎に少量のラベル付きデータでの追加学習（Fine Tuning）を行うというアプローチが取られるようになった。この事前学習で作られたトランスフォーマー型の深層学習モデルが、2018年以降、自然言語処理においてスタンダードになり、改良・拡張を加えたバリエーションが多数生まれている[17)]。特に注目されたのは、2018年にGoogleから発表されたBERT[18)]と、2020年にOpenAIから発表されたGPT-3[19)]である。 BERTを中心とするこれらの技術概要については［新展開・技術トピックス］❶で取り上げ、GPT-3については［注目すべき国内外のプロジェクト］❶で取り上げる。

❹ 知覚・運動系 AI と言語・知識系 AI の統合に関わる動向

パターン処理を中心とした知覚・運動系のAI技術と、記号処理を中心とした言語・知識系のAI技術は、別系統で発展してきた。1950年代後半から1960年代にかけての第1次AIブームと、1980年代の第2次AIブームで扱われたのは、記号処理のAI研究であった。「2.1.1 知覚・運動系のAI技術」で述べたように、第1次AIブームと同時期に、ニューラルネットワークを用いたパターン認識の研究も活発に取り組まれていた。そして、深層学習を中心としたニューラルネットワーク型のAI技術の発展が、2000年代以降の第3次AIブームの主役となった。

このように別系統で発展してきたが、AIとしてパターン処理と記号処理の両面を扱う必要があることは古くから言われていた。また、これまでも2つのタイプのAI技術を組み合わせたシステムは多く見られる。例えば、音声翻訳システムは、音声認識というパターン処理と、機械翻訳という記号処理をつなげたシステムである。また、統計的機械翻訳（SMT）は、記号処理をベースに構成されたシステム中のいくつかのパーツを、機械学習を用いてチューニングしたものである。物理モデルや事前知識モデルを用いたシミュレーションシステムに機械学習を組み合わせたり、機械学習の分類・判定結果を解釈するためにナレッジグラフ

3 アテンション機構には大きく分けると、Self-AttentionとSource-Target-Attentionという2種類がある。アテンションを求める際に、Self-Attentionは対象文中の情報からウェイトを計算し、Source-Target-Attentionは別文中の情報からウェイトを計算する。トランスフォーマーでは、Self-Attention機構をマルチヘッドで動かすことで、複数の言語現象を並列に効率良く学習できるようにしている。

を組み合わせたりといった取り組みも、2つのタイプのAI技術の組み合わせと見ることもできる。

これらに対して、深層学習の発展によって2018年頃から顕在化してきた取り組みは、人間の知能のモデルを意識したパターン処理と記号処理の統合に関する研究である。すなわち、人間の知覚・運動系と言語・知識系の関係や、それらが構成する即応的ループと熟考的ループの情報の流れと対応するような、あるいは、そこからインスパイアされたような、パターン処理と記号処理の統合モデルが検討されている[20), 21)]（その具体例は［新展開・技術トピックス］❷で紹介する）。

このような動きが顕在化したことを示したのが、NeurIPS 2019（The 33rd Conference on Neural Information Processing Systems）でのYoshua Bengioによる「From System 1 Deep Learning to System 2 Deep Learning」と題した招待講演[22)]である。ここで、System 1とSyetem 2は、2002年にノーベル経済学賞を受賞したDaniel Kahnemanのいう直観的な「速い思考」（System 1）と論理的な「遅い思考」（System 2）[23)]のことで、即応的ループと熟考的ループに対応する。 Bengioは、現在の深層学習はSystem 1に相当するが、System 2までカバーするような深層学習へ発展させるのが今後の方向性だと示唆した。さらに、Yoshua Bengio、Geoffrey Hinton、Yann LeCunの3人は、深層学習発展への貢献で2018年度ACM（Association for Computing Machinery）チューリング賞を受賞したが、AAAI 2020（The 34th AAAI Conference on Artificial Intelligence）における同賞記念イベントでは3人の講演に加えてKahnemanを交えたパネル討論も実施され、この方向性が論じられた。一方、日本国内では、これよりも早く、「深層学習の先にあるもの - 記号推論との融合を目指して」と題した東京大学公開シンポジウムが2018年1月と2019年3月[24)]に開催されている。

❺ 学会動向および国際動向

自然言語処理分野の最先端研究は、トップランク国際会議ACL（Annual Meeting of the Association for Computational Linguistics）、EMNLP（Conference on Empirical Methods in Natural Language Processing）、NAACL（North American Chapter of the Association for Computational Linguistics）等で活発に発表されている。 ACL 2020の国別採択論文数は、1位の米国が305件、2位の中国が185件、3位の英国が50件、4位のドイツが44件、日本は5位で24件であった。投稿論文数では中国が米国を上回っており、自然言語処理分野も米中二強になりつつある。

パターン処理と記号処理の統合（知覚・運動系と言語・知識処理の統合）については、AI全般のトップランク国際会議であるAAAI（Association for the Advancement of Artificial Intelligence）やIJCAI（International Joint Conferences on Artificial Intelligence）、あるいは、機械学習分野のNeurIPS（Neural Information Processing Systems）等での発表も目に付く。

米国はGAFA（Google、Amazon、Facebook、Apple）やMicrosoft、IBM等、産業界による先進技術の研究開発・応用が活発である上に、技術政策として、国の安全保障目的も含めた自然言語処理の基礎研究への先行投資が際立っている。もともと米国国立標準技術研究所（NIST）による情報検索・質問応答技術、国防高等研究計画局（DARPA）による情報抽出技術や文章理解、高等研究開発局（ARDA）による質問応答技術の研究開発等、自然言語処理に関わる多くのプロジェクト（コンペティション型ワークショップを含む）が政府予算によって推進されてきた。さらに2018年にDARPAはAI Next Campaignを発表した（［注目すべき国内外のプロジェクト］❷参照）。

中国政府は2017年7月に次世代AI発展計画を発表し、AI産業を強力に推進している。自然言語処理分野ではMicrosoftやBaidu（百度）が目に付く。Microsoftは2014年からチャットボットXiaoice（シャ

オアイス）を公開しており、ソーシャルネットやメッセージングのアプリに導入され、世界中で6.6億人のユーザーがいるという。 BaiduはWebサーチエンジンで実績があるほか、BERTを中国語向けに強化したERNIE[25]も開発している。

欧州は多数の国にまたがることから、欧州フレームワークプログラムの中で、機械翻訳を中心に自然言語処理に継続的に投資を行ってきている。産業界でも、DeepLの機械翻訳サービスが翻訳品質の高さで注目されている。

日本は現在「AI戦略2019」を推進しており、文部科学省による理化学研究所革新知能統合研究センター（AIP）、経済産業省による産業技術総合研究所人工知能研究センター（AIRC）、総務省による情報通信研究機構（NICT）の3つを中核的なAI研究機関と位置付けている。自然言語処理については、NICTが機械翻訳・音声翻訳を中心に取り組んでおり、その実用化でも実績があるほか、AIPで基礎研究を推進している。パターン処理と記号処理を統合した次世代AI研究については、新エネルギー・産業技術総合開発機構（NEDO）による「次世代人工知能・ロボット中核技術開発」事業においてAIRCが中心的に取り組んでいるほか、文部科学省の2020年度戦略目標「信頼されるAI」とそれを受けた科学技術振興機構（JST）の戦略的創造研究推進事業（CREST等）でも基礎研究面が強化された。

（4）注目動向

［新展開・技術トピックス］

❶ トランスフォーマー事前学習モデル

［研究開発の動向］❸で述べたように、現在（2020年末時点）、トランスフォーマー型のニューラルネットワークで、ラベルなしの大量テキストデータを用いた事前学習を行うことで、汎用性の高い言語モデルを構築するアプローチが、自然言語処理のスタンダードになっている。特に、2018年にGoogleから発表されたBERT[18]がよく知られているが、これに類するモデルや、これを改良・拡張したモデルが2018年以降、種々発表されている。2018年にはBERTの前にOpenAIのGPT[26]、アレンAI研究所のELMo[27]が発表されていた。 GPTはLeft-to-Right単方向のトランスフォーマー、ELMoはLeft-to-RigtとRight-to-LeftのLSTMの連結、BERTは双方向のトランスフォーマーである。

ラベルなしの大量テキストデータを用いた事前学習、つまり、自己教師あり学習を行うために、BERTではMLM（Masked Language Model）が導入された[4]。MLMはもとのテキストに対して複数個所をマスクし（隠し）、穴埋め問題のようにマスク箇所を当てるというタスクを、大量テキストデータで訓練するというものである。もとのテキストがあって穴埋め問題の答えは分かるので、このタスクは実質的に教師あり学習として訓練できる。ニューラルネット自然言語処理の初期、Word2Vecで用いられたSkip-Gram法やCBOW法による分散表現の獲得では、多義語の意味を分離できなかった。これに対して、MLMでは文脈を考慮して多義語を区別した分散表現を獲得できる。自然言語処理のベンチマークとしてGLUE（General Language Understanding Evaluation）、SQuAD（Stanford Question Answering Dataset）等があるが、BERTの登場によって、それらの多くのタスクにおいて最高スコアが更新された。

その後、BERTを改良・拡張したモデルが次々に考案され、2019年・2020年で10種類以上発表され

4 より詳細には、MLMとともに、NSP（Next Sentence Prediction）がBERTに導入された。 NSPは2つの文が連続する文かどうかを判定するタスクを学習するものである。

た[17]。ベンチマークの最高スコアも次々に更新されている。XLNet[28]は、MLMを単語の並べ替えによって改良し、BERTのスコアを上回った。しかし、必要な計算資源の規模拡大が著しく、1回の学習に要する費用は、クラウド利用料に換算すると、BERTの場合で6.9千USドル（日本円で約72万円）、XLNetの場合で61千USドル（日本円で約640万円）にも上る[29]。また、RoBERTa[30]はBERTのMLMをダイナミックにする改良を加えて性能を改善した。ALBERT[31]はBERTを軽量化したモデル、ERNIE[25]は中国語向けにBERTを改良したモデルであり、UniLM[32]は複数の言語モデルを使い分ける。ViBERT[33]、VL-BERT[34]、UNITER[35]等はBERTをマルチモーダルに拡張したモデルである。

❷ 深層学習の発展・拡張による知能のモデル化

［研究開発の動向］❹で述べたように、パターン処理と記号処理を比較的疎な形で組み合わせたシステム化は以前から行われてきたが、深層学習が発展し、自然言語処理やナレッジグラフを用いた処理のような記号レベルの処理も深層学習によって再構成されるようになった。これにより、パターン処理と記号処理を統一的な考え方で統合する（共通の枠組み上に融合する）可能性が見えてきた。従来よりも密な融合形としては、深層学習ベースの言語モデル中に外部知識（ナレッジグラフ、意味ネットワーク）を組み込むアプローチ（例えばSenseBERT[36]）や、パターン処理結果を命題論理表現に変換してソルバーで推論できるようにするアプローチ（例えばFOSAE[37]）等もあるが、人間の知能のモデルを意識した融合形が特に注目される。もともと深層学習や強化学習は、人間の知覚・運動系に類する学習モデルであり、その拡張として言語・知識系までカバーしようとするのは、自然な発展の方向である。「2.1.7 計算脳科学」や「2.1.8 認知発達ロボティクス」の研究領域で人間の知能に関する研究が進展し、それに基づくニューロシンボリックAI（Neuro-Symbolic AI）、神経科学（脳科学）にインスパイアされたAI（Neuroscience-Inspired AI）、記号創発ロボティクス（Symbol Emergence in Robotics）といったコンセプトでの研究の流れが形作られている。

例えば、NS-CL（Neuro-Symbolic Concept Learner）は、画像や環境・空間の認識と質問応答という2系統を持ち、画像・空間系統は教師あり学習、質問応答系統は強化学習を用い、2系統のマルチタスクのカリキュラム学習を通して、視覚的概念・単語・意味解析等を学習していく[38]。また、知能の2階建てモデルは、1階部分が深層学習をベースとした知覚・運動系で、その外界とのインタラクションを通して作られた世界モデルを介して、2階部分のトランスフォーマー的な言語・知識系が動くというものである[21]。記号創発ロボティクスでは、外界とのインタラクションを通して言語が創発的に形成されることを、確率的生成モデルをベースにモデル化することを試みている[21]。

❸ マルチモーダル・マルチタスク学習

意味の分散表現のベクトル形式は、テキストデータだけでなく、画像・映像・音声等の異なるデータタイプの入力に対しても用いることが可能である。また、近年、グラフニューラルネット等の手法が発展し、ナレッジグラフにも分散表現が用いられるようになった。こうしたことから、マルチモーダル処理を共通的なニューラルネットワーク構造で行うことが容易になっている。さらに、それらの入力に対して行うタスクも共通構造をベースに組み上げることが可能である。実際に、そのような共通構造をマルチモーダル・マルチタスクで訓練することで、個別にニューラルネットワークを訓練した場合よりも精度・効率が向上することが報告されている[39]。

このような効果をもたらす仕組みは、前述したBERTのマルチモーダル拡張版（ViBERT[33]、VL-

BERT[34]、UNITER[35]）にも取り入れられている。これらの中で最も高い性能を示しているUNITERでは、テキストと画像のそれぞれに対する一部を隠したモデリング（Masked Modeling）だけでなく、画像とテキストの対応付け、単語と画像領域の対応付け等、マルチモーダル型のタスクも加えて、計4種類の事前学習を用いている。

また、このようなマルチモーダルモデルを用いることで、画像・動画に対する説明文（キャプション）の生成、画像・動画の内容に関する質問・回答、テキストによる画像・動画検索、画像・動画の説明文から画像中の参照箇所の抽出といった応用が可能になる。

［注目すべき国内外のプロジェクト］

❶ Open AI GPT-3

OpenAIは2018年にトランスフォーマー型の事前学習モデルGPT[26]を発表した後、2019年にその第2世代GPT-2、さらに2020年に第3世代のGPT-3[19]を発表した。GPT-3は極めて巨大な言語モデルで、モデルのパラメーター数は1750億個（BERTの500倍以上、GPT-2の100倍以上）、事前学習に用いられたデータ量は45TB（BERTの約3000倍）、事前学習に要した計算量は3640 petaflop/s-day（毎秒1ペタ回のニューラルネット演算を10年近く実行した計算量に相当）にも及んだ。この超大規模の事前学習を行った結果、少数の例示をもとに文章を生成できることが示された（従来のタスク毎の追加学習に対してFew Shot学習と呼ばれる）。GPT-3のAPI（Application Programming Interface）を利用したさまざまな応用試作が行われており、例示文に続けてブログや小説を生成したり、簡単な機能説明文からプログラムコードやHTMLコードを生成したり、さまざまな質問文に対して回答を生成したりといった事例が次々と発表されている。また、同種のモデルを画像とテキストのペアで訓練したDALL・Eというモデル（パラメーター規模は120億個）を用いると、自然言語を与えるとそれに対応する画像・イラストを生成することができる。

GPT-3が生成する文章には、人間が書いたかのような自然さがあり、GPT-3のAPIを用いて掲示板に自動投稿されていた文章が、1週間、人間が書いたものでないと見破られなかったということである。このようにGPT-3の機能・性能は衝撃的なものであるが、その一方、フェイクニュース生成器になり得ることや、訓練データ中に存在した差別・偏見が生成文に反映されること等、GPT-3が持つ危険性・リスクも認識されている。また、質問入力に対するGPT-3の回答出力を分析すると、物理法則・数学式・常識に反する等、意味を理解して文章を生成しているわけではないことが分かる。

❷ DARPA AI Next

米国DARPAは2018年に、AI研究に20億ドル以上の大型投資を実施するというAI Next Campaignを発表した。この発表では、AIの発展を、専門家の知識を抽出・活用するHandcrafted Knowledgeを「第1の波」、ビッグデータから知見を導く深層学習に代表されるStatistical Learningを「第2の波」に続く「第3の波」として、文脈を理解して推論するContextual Reasoningを挙げた。これによって、人間が把握・理解・行動する以上の速度でデータを生成・処理し、安全かつ高度に自律的な自動化システムを可能にしつつ、人間の意思決定を支援し、人間と機械の共生を促進することを狙っている。

DARPAでは、AI Nextを掲げつつ、XAI（Explainable AI）、MCS（Machine Common Sense）、L2M（Lifelong Learning Machines）、SemaFor（Semantic Forensics）をはじめ、言語・知識系AIの高度化や、知覚・運動系AIとの統合につながるようなさまざまな先進プロジェクトを推進している。

（5）科学技術的課題

❶ 真の意味理解・常識推論

BERTをはじめさまざまな言語モデルが提案されているが、それらの性能評価には共通のベンチマークが用いられる。その代表的なタスクとして、テキストを読んだ上で、その内容に関する質問に答える文書読解タスクがある。文書読解のベンチマークとしてよく使われているのがSQuADである。また、単一タスクでなく多数のタスクを評価できるベンチマークのセットとして、GLUEもよく使われている。 GLUEには含意関係判定、同義性判定、質問・回答解析、肯定的か否定的かの感情解析、文法チェック等の約10種類のタスクが用意されている。

言語モデルの改良により、これらのベンチマークでのスコアが人間の平均的スコアを上回ったという報告も出てきている。しかし、文章の意味を理解しなくても、統計的な傾向を捉えれば正解を当てられるような問題が多かったため、見かけ上、高いスコアが得られただけで、本当に意味を理解しないと当たらないような問題に対してはスコアが低くなることが指摘されている[40), 41)]。自然言語処理研究の黎明期からその難しさも含めて認識されている常識推論に向けて、含意関係認識やストーリー予測等のサブタスクの設定、敵対的サンプル（Adversarial Examples）等も取り入れたベンチマークの構築が求められる[42)]。その試みの一つとして、意味を理解しないと正答が難しい問題を多く含む文書読解タスクのベンチマークDROP[43)]が公開された。 SQuADでは既にトップスコアが90%を超えたが、DROPではBERTのスコアが32.7%、一方で人間のスコアは96.4%と、機械にとって難しいベンチマークとなっている。

2020年に登場し、衝撃的な機能・性能を示したGPT-3においても、意味理解・常識推論は課題とされており、依然として難しく重要な課題である。

❷ 言語・知識系と知覚・運動系を統一的に扱う次世代 AI アーキテクチャー

［新展開・技術トピックス］❷で述べたように、言語・知識系AIと知覚・運動系AIの統合を、深層学習を発展させた枠組みで実現する取り組みが、次世代のAIアーキテクチャーとして目指されている[20), 21), 24), 44)]。この新たな枠組みによって、a.学習に大量の教師データや計算資源が必要であること、b.学習範囲外の状況に弱く、実世界状況への臨機応変な対応ができないこと、c.パターン処理は強いが、意味理解・説明等の高次処理はできていないこと、といった現在の深層学習の抱える問題の解決が期待される。これらの問題の解決には、古くからAIの基本問題として掲げられている記号接地問題やフレーム問題も関わっており、これを再定義して段階的に解決していくようなアプローチが必要になると思われる。

GPT-3は、トランスフォーマー型モデルによる現在のAIアーキテクチャーでも、大規模な訓練データを用意して大規模なモデルを作ることで、人間と区別するのが難しいほどの生成能力を実現できることを示した。このようなある種力任せのアプローチと、新たな次世代AIアーキテクチャーとの対比から得られる知見も重要なものになると思われる。

❸ マルチモーダル処理・実世界連結の応用開発

［新展開・技術トピックス］❸で述べたように、トランスフォーマー型の事前学習モデルがマルチモーダルに拡張され、性能を向上させている。今後、これを用いたマルチモーダルやクロスモーダルの応用が広がっていくものと思われる。例えば、画像・動画に対する説明文（キャプション）の生成、画像・動画の内容に関する質問・回答、テキストによる画像・動画検索、画像・動画の説明文から画像・動画中の参照箇所の抽出といった応用が試みられている[45)]。その発展として、ロボットと組み合わせた実世界連動型タ

スクの実現も見込まれる。具体的な事例として、Preferred NetworksがCEATEC JAPAN 2018（2018年10月）でデモンストレーション展示を行った「全自動お片付けロボットシステム」がある。このシステムでは、部屋の中に散らかったさまざまな物について、片付けロボットに対して「それは赤い箱に片づけて」というような物体移動の命令を言語指示で行うことができる[46)]。また、認知発達・記号創発ロボティクスのアプローチとして、ロボットが実世界とのマルチモーダルなインタラクションを通して、場所概念や語彙を獲得していくプロセスの試作も行われている[21),44)]。

（6）その他の課題

❶ 新たな研究課題のための訓練データやベンチマークデータの戦略的構築

かつては形態素解析・構文解析等の自然言語処理のサブタスク別の精度評価が、技術改良における主たる目標になっていた。しかし、ニューラルネット自然言語処理では応用毎のEnd-to-End最適化のアプローチがとられるため、サブタスクに注力することのウェイトが下がってきた。このような状況で、近年は、GLUEのように問題毎のベンチマークタスクを複数用意して、言語モデルの汎用性あるいは特定用途の有用性を評価するようになった。

同時に、AI分野では共通タスク・共通データセットでのコンペティション型ワークショップが盛んに実施されてきた。特に米国NISTが自然言語処理・情報検索領域でこれを推進してきたことは先駆的で、この分野の基礎研究の強化を大きく牽引した。このような活動の推進においては、タスク設定に関わる目利き人材がキーになる。取り組むことに大きな価値があり、しかも無理難題ではなく挑戦意欲をかき立てるようなタスク設定の匙加減を適切にでき、コミュニティーをリードできるような人材が求められる。

このようなベンチマークやコンペティションのタスク設定・データセット構築は、米国がリードし、中心的に貢献してきたが、日本においても、データセット構築とコンペティション型ワークショップ運営に20年以上取り組んでいる国立情報学研究所のNTCIR（NII Testbeds and Community for Information access Research）プロジェクト等の実績がある。今後、言語・知識系と知覚・運動系の統合AIや、マルチモーダルAI・実世界連結応用等の新しい研究課題に取り組んでいくにあたっては、新たなタスク設定・データセット構築が重要である。日本としての戦略的取り組みが期待される。

また、テキストデータの場合、国際的な競争や協力の面では、英語テキストが主にならざるを得ないし、デジタルデータの流通量という面でも英語テキストが優位である。これに対して、そもそも集め得るデータ量が格段に少ない国の言語（低資源言語と呼ばれる）に対する取り組みや、日本語固有の問題への取り組みをどう進めるかという点も検討を要する。

❷ 大規模コンピューティング基盤の共同利用施設とその継続的強化・整備

BERT、XLNet、GPT-3の言語モデルの事前学習に要する計算資源の規模について前述したが、最新の言語モデル開発には極めて大量の計算資源を必要とし、その実行環境は大学の一研究室で確保できる規模ではなくなっている。大規模コンピューティング基盤の共同利用施設が不可欠であり、産業技術総合研究所のAI橋渡しクラウドABCI（AI Bridging Cloud Infrastructure）や理化学研究所のスーパーコンピューター「富岳」がこの役割を担っている。特にABCIでは、BERTの事前学習済みモデルを共同利用できるように公開している。このような共同利用施設・計算資源の継続的な強化・整備が極めて重要である。

❸ 人間・社会面の深い理解・考察に基づく取り組み・人材育成

GPT-3が人間と区別できないような文章を生成できることの危険性が指摘されているように、この研究分野から生み出される技術がもたらす社会的影響が増大している。また、今後の発展においては、人間の知能のメカニズムに学ぶという面が強くなりつつある。 AI全般の倫理的・法的・社会的課題（Ethical, Legal and Social Issues：ELSI）面については「2.1.9 社会におけるAI」にて論じるが、この分野の技術の社会的影響や知能そのものに関する深い理解・考察とともに研究開発を進める必要があろう。人間の知能の情報処理メカニズムの理解という面では、「2.1.7 計算脳科学」や「2.1.8 認知発達ロボティクス」は日本が実績のある研究分野であり、その研究成果をこの分野の発展・強みに活かしていきたい。

1980年～2000年頃、ルールベース方式が主流だった時代は、カナ漢字変換、機械翻訳、サーチエンジン等をターゲットとして、電気系大手企業の各社が自然言語処理の研究開発に力を入れていた。しかし、機械学習を用いる方式では、大量のテキストデータを使うことが研究開発に不可欠で、多数のユーザーを抱えるインターネットサービスを運営している企業において、自然言語処理への取り組みが活発になってきた。特にGAFAは保有するデータ量やそれを処理する計算機パワーが圧倒的で、データ収集や実験が行いやすいことは研究者に魅力的である。 AI分野の人材争奪戦が国際的に激しくなる中で、AI人材の教育・育成、幅広い人材の獲得や引き留めのための施策も重要である。

（7）国際比較

国・地域	フェーズ	現状	トレンド	各国の状況、評価の際に参考にした根拠など
日本	基礎研究	○	↗	「AI戦略2019」を推進しており、理研AIP、産総研AIRC、NICTの3つを中核的なAI研究機関と位置付けている。自然言語処理については、NICTが機械翻訳・音声翻訳を中心に取り組んでおり、その実用化でも実績があるほか、AIPで基礎研究を推進している。パターン処理と記号処理を統合した次世代AI研究については、NEDO「次世代人工知能・ロボット中核技術開発」事業においてAIRCが中心的に取り組んでいるほか、文科省の2020年度戦略目標「信頼されるAI」とそれを受けたJST戦略的創造研究推進事業でも基礎研究面が強化された。言語処理学会に活気があり、ACL・EMNLP等の国際的トップカンファレンスでも一定数の発表がなされている。言語資源や研究利用可能なコンテンツの整備が不足しているとともに、言語モデルの超大規模化に対して計算機環境面で劣勢である。
	応用研究・開発	○	→	NICTが音声翻訳、大規模情報分析、災害関連情報分析等で実用性の高い研究成果をリリースしてきた。 NEC、NTTドコモ、日本IBM、Softbank、トヨタ等の民間企業でも、自然言語処理技術に基づく製品化・事業化が行われてきた。
米国	基礎研究	◎	→	GAFA、Microsoft、IBM等、産業界による先進技術の研究開発・応用が活発である上に、技術政策として、国の安全保障目的も含めた自然言語処理の基礎研究への先行投資が際立っている。 ACL、EMNLP等の有力な国際会議等での発表の多くは米国発である。もともとNISTによる情報検索・質問応答、DARPAによる情報抽出・文章理解、ARDAによる質問応答等、自然言語処理技術に関わる多くのプロジェクトやコンペティションが政府予算によって推進されてきた。さらに2018年にDARPAはAI Next Campaignを発表した。

	応用研究・開発	◎	↗	上述した基礎研究が大学教員の移籍等によって、比較的短期間にGoogle、Microsoft、IBM、Facebook、Amazon等の応用研究・開発に回るサイクルが確立している。これらの企業の研究開発への投資も大きく、また、ベンチャーによる取り組みも活発である。
欧州	基礎研究	○	→	欧州は多数の国にまたがることから、欧州フレームワークプログラムの中で、機械翻訳を中心に自然言語処理に継続的に投資を行ってきているが、突出した研究は少ない。
	応用研究・開発	○	→	グローバル企業の研究所が存在し、一定のアクテビティーはあるが、米国主導による産業化の側面が強い。産業界では、DeepLの機械翻訳サービスが翻訳品質の高さで注目されている。
中国	基礎研究	◎	↗	北京大学・清華大学等の有力大学やMicrosoft Research Asia、Baidu等の民間企業の研究所を中心に基礎研究が進められている。 ACL等のトップ国際会議でも論文採択数は米国に次いで中国が2位である。
	応用研究・開発	◎	↗	中国政府は2017年7月に次世代AI発展計画を発表し、AI産業を強力に推進している。自然言語処理分野ではMicrosoftやBaidu（百度）が目に付く。Microsoftは2014年からチャットボットXiaoice（シャオアイス）を公開しており、ソーシャルネットやメッセージングのアプリに導入され、世界中で6.6億人のユーザーがいるという。 BaiduはWebサーチエンジンで実績があるほか、BERTを中国語向けに強化したERNIEも開発している。
韓国	基礎研究	△	→	KAIST、ETRI、KISTI等の有力大学、国研を中心に基礎研究が進められている。ただ、ACL等のトップコンファレンスでの韓国発の発表は減少している。
	応用研究・開発	○	↗	政府がNaver、Samsung、SK Telecom、Hyundai等とArtificial Intelligence Research Instituteを設立予定。 Naver等によるチャットボットや機械翻訳のサービスの開始が相次ぐ。

（註1）フェーズ

基礎研究：大学・国研などでの基礎研究の範囲

応用研究・開発：技術開発（プロトタイプの開発含む）の範囲

（註2）現状　※日本の現状を基準にした評価ではなく、CRDSの調査・見解による評価

◎：特に顕著な活動・成果が見えている　　○：顕著な活動・成果が見えている

△：顕著な活動・成果が見えていない　　×：特筆すべき活動・成果が見えていない

（註3）トレンド　※ここ1～2年の研究開発水準の変化

↗：上昇傾向、→：現状維持、↘：下降傾向

参考文献

1）Christopher D. Manning, “Computational linguistics and deep learning”, *Computational Linguistics* Vol.41, No.4（2015）, pp.701-707. DOI：10.1162/COLI_a_00239

2）Minh-Thang Luong, Hieu Pham and Christopher D. Manning, “Effective Approaches to Attention-based Neural Machine Translation”, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*（EMNLP 2015; Lisbon, Portugal, September 17–21, 2015）, pp.1412-1421.

3）久保陽太郎,「ニューラルネットワークによる音声認識の進展」,『人工知能』(人工知能学会誌）31巻2号（2016年3月）, pp.180-188.

4）中澤敏明,「機械翻訳の新しいパラダイム：ニューラル機械翻訳の原理」,『情報管理』66巻5号（2017年8月）, pp.299-306. DOI：10.1241/johokanri.60.299
5）Ilya Sutskever, Oriol Vinyals and Quoc V. Le, “Sequence to Sequence Learning with Neural Networks”, arXiv：1409.3215（2014）.
6）岡崎直観,「言語処理における分散表現学習のフロンティア」,『人工知能』(人工知能学会誌）31巻2号（2016年3月）, pp.189-201.
7）Maximilian Nickel, et al., “A Review of Relational Machine Learning for Knowledge Graphs”, *Proceedings of IEEE* Vol. 104, No. 1（2016）, pp.11-33.
8）Tomas Mikolov, et al., “Distributed Representations of Words and Phrases and Their Compositionality”, *Proceedings of 27th Annual Conference on Neural Information Processing Systems*（NIPS 2013; Nevada, United States, December 5-8, 2013）, pp.3111-3119.
9）坪井祐太・海野裕也・鈴木潤,『深層学習による自然言語処理』(講談社,機械学習プロフェッショナルシリーズ,2017年）.
10）情報処理推進機構AI白書編集委員会（編）,「自然言語を中心とする記号処理」,『AI白書2017』（KADOKAWA,2017年）, pp.64-78（1.4節）.
11）Anand Rajaraman and Jeffrey David Ullman, *Mining of Massive Datasets*（Cambridge University Press, 2012）.（邦訳：岩野和生・浦本直彦訳,『大規模データのマイニング』,共立出版,2014年）
12）大塚裕子・乾孝司・奥村学,『意見分析エンジン―計算言語学と社会学の接点』(コロナ社,2007年）.
13）Special Issue：“This is Watson”, *IBM Journal of Research and Development* Vol. 56 issue 3-4（May-June 2012）.
14）Dzmitry Bahdanau, Kyunghyun Cho and Yoshua Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate”, *Proceedings of the 3rd International Conference on Learning Representations*（ICLR 2015; San Diego, CA, USA, May 7-9, 2015）.
15）西田京介・斉藤いつみ,「深層学習におけるアテンション技術の最新動向」,『電子情報通信学会誌』101巻6号（2018年）, pp. 591-596.
16）Ashish Vaswani, et al., “Attention Is All You Need”, *Proceedings of the 31st Conference on Neural Information Processing Systems*（NIPS 2017; Long Beach, CA, USA, December 4-9, 2017）.
17）丹羽彩奈,「NeurIPS2019における自然言語処理：Attentional Neural Network Modelsの進展-TransformerとBERTを中心に-」,『人工知能学会第78回人工知能セミナー：AIトレンド・トップカンファレンス報告会』(2020年3月）.
18）Jacob Devlin, et al., “BERT： Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv：1810.04805（2018）.
19）Tom Brown, et al., “Language Models are Few-Shot Learners”, *Proceedings of the 34th Conference on Neural Information Processing Systems*（NeurIPS 2020; December 6-12, 2020）.
20）科学技術振興機構 研究開発戦略センター,「戦略プロポーザル：第4世代AIの研究開発―深層学習と知識・記号推論の融合―」,CRDS-FY2019-SP-08（2020年3月）.

21）科学技術振興機構 研究開発戦略センター,「JSAI2020企画セッション報告書（2020年6月9日開催）：次世代AI研究開発―さらなる進化に向けて―」,CRDS-FY2020-XR-02（2020年10月）.
22）Yoshua Bengio, “From System 1 Deep Learning to System 2 Deep Learning”, *Invited Talk in the 33rd Conference on Neural Information Processing Systems*（NeurIPS 2019; Vancouver, Canada, December 8-14, 2019）. https://slideslive.com/38922304/from-system-1-deep-learning-to-system-2-deep-learning（accessed 2021-01-30）
23）Kahneman, D., *Thinking, Fast and Slow, Farrar*（Straus and Giroux, 2011）.（邦訳：村井章子訳,『ファスト＆スロー：あなたの意思はどのように決まるか?』,早川書房,2014年）
24）東大TV,「深層学習の先にあるもの - 記号推論との融合を目指して（2）」（2019年3月5日）. https://todai.tv/contents-list/2018FY/beyond_deep_learning（accessed 2021-01-30）
25）Yu Sun, et al., “ERNIE：Enhanced Representation through Knowledge Integration”, arXiv：1904.09223（2019）.
26）Alec Radford, et al., “Improving language understanding with unsupervised learning”, OpenAI Technical report, OpenAI Blog（11 June 2018）. https://blog.openai.com/language-unsupervised/（accessed 2021-01-30）
27）Matthew E. Peters, et al., “Deep contextualized word representations”, arXiv：1802.05365（2018）.
28）Zhilin Yang, et al., “XLNet：Generalized Autoregressive Pretraining for Language Understanding”, arXiv：1906.08237（2019）.
29）“The Staggering Cost of Training SOTA AI Models”, Synced（June 17, 2019）. https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/（accessed 2021-01-30）
30）Yinhan Liu, et al., “RoBERTa：A Robustly Optimized BERT Pretraining Approach”, arXiv：1907.11692（2019）.
31）Zhenzhong Lan, et al., “ALBERT：A Lite BERT for Self-supervised Learning of Language Representations”, arXiv：1909.11942（2019）.
32）Li Dong, et al., “Unified Language Model Pre-training for Natural Language Understanding and Generation”, arXiv：1905.03197（2019）.
33）Jiasen Lu, et al., “ViLBERT：Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks”, arXiv：1908.02265（2019）.
34）Weijie Su, et al., “VL-BERT：Pre-training of Generic Visual-Linguistic Representations”, arXiv：1908.08530（2019）.
35）Yen-Chun Chen, et al., “UNITER：UNiversal Image-TExt Representation Learning”, arXiv：1909.11740（2019）.
36）Yoav Levine, et al., “SenseBERT：Driving Some Sense into BERT”, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*（ACL 2020; July 5-10, 2020）.
37）Masataro Asai, “Unsupervised Grounding of Plannable First-Order Logic Representation from Images”, *Proceedings of the 29th International Conference on Automated Planning and Scheduling*（ICAPS 2019; Berkeley, CA, USA, July 10-15, 2019）.

38）Jiayuan Mao, et al., “The Neuro-Symbolic Concept Learner：Interpreting Scenes, Words, and Sentences From Natural Supervision”, *Proceedings of the 7th International Conference on Learning Representations*（ICLR 2019; New Orleans, USA, May 6-9, 2019）.
39）Lukasz Kaiser, et al., “One Model To Learn Them All”, arXiv：1706.05137（2017）.
40）Robin Jia and Percy Liang, “Adversarial Examples for Evaluating Reading Comprehension Systems”, arXiv：1707.07328（2017）.
41）Saku Sugawara, et al., “What Makes Reading Comprehension Questions Easier?”, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*（EMNLP 2018; Brussels, Belgium, October 31-November 4, 2018）, pp. 4208-4219.
42）井之上直也,「言語データからの知識獲得と言語処理への応用」,『人工知能』(人工知能学会誌）33巻3号（2018年5月）, pp. 337-344.
43）Dheeru Dua, et al., “DROP：A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs”, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*（NAACL 2019; Minneapolis, Minnesota, USA, June 2-7, 2019）.
44）科学技術振興機構 研究開発戦略センター,「科学技術未来戦略ワークショップ報告書：深層学習と知識・記号推論の融合によるAI 基盤技術の発展（2020年1月30日開催）」, CRDS-FY2019-WR-08（2020年3月）.
45）牛久祥孝,「画像に関連した言語生成の取組み」,『人工知能』(人工知能学会誌）34巻4号（2019年7月）, pp. 483-491.
46）羽鳥潤・他,「実世界におけるインタラクティブな物体指示」,『言語処理学会第24回年次大会発表論文集』（2018年3月）, pp.943-946.