

言語の機械的な処理と人間の認知的な処理の類似性について検討する

大谷直輝

1. 認知言語学における実証的な研究を求めて

認知言語学は人間の心（認知能力）の観点から、言語の構造や機能を明らかにしようとする言語学のアプローチである。1980年代に起こり、初期の頃は、認知文法論、認知意味論、概念メタファー論等の研究が行われていたが、現在は学問としての体系性がまし、認知音韻論、認知形態論、認知語用論、構文文法論、認知類型論等の研究が広く展開されている。認知言語学は、言語と認知の関係について、2つの基本的な立場をとる。第一に、人間の認知能力が言語の構造や機能に影響を与えると主張する。つまり、五感を通して世界を認識する人間の知覚や概念化における特徴が言語の構造や機能に反映すると考える。第二に、人間の言語知識は言語使用の中から創発すると主張する。言語使用において繰り返して用いられるパターンの中に類似性が見つかる中で、パターンの抽象化や切り分けが行われることで、文法や語に関する知識が習得されると考える。

認知言語学は実証的な研究を行うのが難しい学問である。認知言語学は、その名が示す通り、言語の分析をする際に、言語学内部の事象を論じるだけでなく、認知に関する様々な側面を考慮した分析を行う必要がある。しかし、認知言語学が学問としての形を見せ始めてからすでに40年がたち、多くの理論的な示唆が積み重なる一方で、認知言語学のおもな方法論は、伝統的な言語学の方法論であるミニマルペアの比較、初歩的なコーパス分析、インフォーマント調査等にとどまっており、人間の心という捉えがたい対象の特性を捉えるための学際的な方法論が普及してるとは言えない。認知言語学の現状を見ると、量的転回(2009～)、社会的転回(2010～)という二つの方法論的な転回を経て、実証性に関する関心が高まる一方で、Dąbrowska and Divjak (2019)が指摘するように、実証的な研究パラダイムとなるには、1) 仮説の検証における統計的と行動的な手法の導入、2) 認知科学(心理学、脳科学、自然言語処理)を用いた学際的な実証的研究、3) 言語の社会性を重視した研究(談話・文脈研究、変化・変異・変種、習得と認知社会能力)という点における、進展が求められる。

本稿では、認知言語学が、親和性の高い分野との学際的な研究が模索する中で、認知言語学が解明を目指す「言語使用の中から創発する知識」の内実を実証的に検証する方法論として、自然言語処理において近年飛躍的な発展を見せている深層学習が様々な用途に利用できる可能性に注目する。使用基盤モデル(the usage-based model)という共通の理念を有する自然言語処理の研究手法を取り入れることで、実証的な研究に必要な、1) データの収集、2) 記述・分類・分析、3) 仮説構築、4) 検証、5) フィードバックという研究のサイクルを確立できる可能性を検討すると同時に、「言語知識とはどのようなものでどのように習得されるのか」という言語学の根源的なテーマを追求する際に、深層学習のアプローチから多くの有益な貢献がなされると主張する。

2. 認知言語学と自然言語処理の共通点

認知言語学の実証的な研究を行う際に、自然言語処理の深層学習を取り入れることが有益であると考えられる理由に、両者は使用基盤モデルをはじめとした理念を共有している点が挙げられる。使用基盤モデルでは、言語使用において繰り返し用いられるパターンからカテゴリー化やスキーマ化が起こり、文法や語彙を含めた言語知識が創発すると考える。同様に、深層学習でも、発話された文をインプットとして、言語が使用される文脈の中から創発する言語知識のありようを追求する。さらに、両者は学際性の高い研究アプローチという共通点もある。認知言語学では、前提として、1) 言語のあらゆる側面に対して、一般的な認知能力の観点から特徴づけを行うべきという「一般化の責任 (Generalization Commitment)」と、2) 心理学や脳科学等の心を扱う学問の知見と矛盾をしない一般化を行うべきという、「認知的責任 (Cognitive Commitment)」を定めており、言語学外部にも開かれた学問を目指している (Lakoff 1990)。一方で、深層学習においても、言語を理解・産出するための言語モデル(ニューラル・ネット)は脳を模倣して作られているという特徴が見られるように、他分野の知見を援用している。

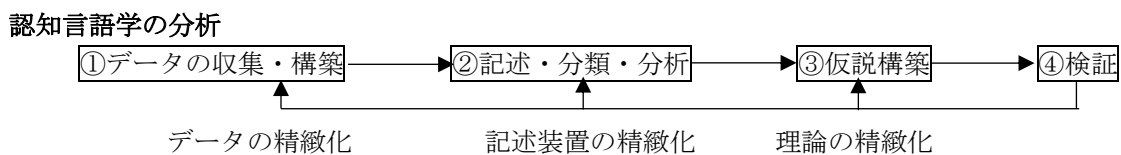
以上の点に注目すると、自然言語処理の技術から導かれる言語知識や処理の過程は、使用基盤モデルが想定する「言語使用の中から創発する知識」や「言語の産出や理解における認知的な処理」の近似値として捉えられる可能性が見えてくる。つまり、認知言語学と深層学習は、実際に発話された文から抽出される言語知識を追求するという点が共通しているだけでなく、深層学習に用いられる言語モデル(BERTなど)は脳を模倣したものであるため、認知言語学の分析によって想定される言語知識と、BERT等の言語モデルから実際に抽出されるアウトプットが類似する場合、1) 認知言語学が提唱する仮説や言語知識が、自然言語処理の観点から間接的に裏づけられると同時に、2) 深層モデルの処理過程における様々な特徴に注目することで、人間が

言語を産出したり理解したりする際の処理過程に関する傾向に対する示唆が得られるⁱ。

その他にも、深層学習をはじめとした、自然言語処理の技術を認知言語学の分析に用いる利点には以下のようなものがある。第一に、言語学の知見だけでは捉え難い、認知能力を可視化できる可能性が考えられる。認知言語学では、人間が持つ一般的な認知能力（カテゴリー化、スキーマ化、前景化など）や学習機構とは矛盾しない理論の構築が求められる。そのため、心理学や脳科学などとの学際的な研究を展開する必要があるが、カテゴリーの内部構造、知識のネットワーク、イメージスキーマ等を言語学者が分析した際に、同じような構造が、自然言語処理の技術を用いて作成できたり再現できたりすることで、言語分析の実証性が増すと考えられる。第二に、自然言語処理の技術を用いることで、言語学者が量的な研究を行う際に必要となる、データの構築やデータ処理を行う際の効率性が大幅に進展する。

3. 実証的な認知言語学の研究サイクルの構築と自然言語処理の役割

さらに、自然言語処理の手法を認知言語学へ導入することで、認知言語学だけでは確立が難しい、実証的な研究のサイクルの構築が促進される可能性がある。



上記の図が示す研究のサイクルにおいて、現在の認知言語学では、おもに、②の認知的な視点からの記述・分類・分析と、③の仮説の構築に関する研究が盛んである。一方で、仮説を構築した後に必要となる、④の仮説の検証については、方法論的な制約もあり発展途上である。また、①のデータの収集・構築に関しても、大規模コーパスの整備が進む一方で、データマイニングで用いられるような実用的なデータの構築が進んでいるとは言い難い。このような現状において、自然言語処理の手法は、④の仮説の検証や、①のデータの収集や構築に大きな貢献を行うことが期待されるⁱⁱ。さらに、仮説の検証を行うことで、研究サイクルの各段階に対するフィードバックが可能となり、データ収集・構築、記述・分類・分析、理論構築といった言語分析の各段階における精緻化が可能となる。

4. 課題となる問題点

以上のように、実際に使用された言語から構築される言語知識を追い求めるという共通理念を持つ認知言語学と深層学習では、学際的な研究の進展が期待される一方で、人間の実際の脳における処理過程を言語モデルに見立てるという方法論に対する懸念も考えられる。よく知られた指摘に、人工知能を哲学的に論じた「中国語の教室」という John Searle によって行われた思考実験がある (Searle 1980)。この思考実験では、ある個別言語に関する知識が無くても、外から見るとその言語の話者とコミュニケーションが成立している状況が示されている。この思考実験によって示されるのは、強い (=人間の脳を模倣した) 人工知能は検証不可能であるという含意である。これまで、永田・大谷・高村・川崎 (2022) 等で、深層学習によるカテゴリー化と人間によるカテゴリー化の共通性が示されているが、必要なことは、このような学際的な研究を積み重ねていくことで、思考実験としての中国語の教室の含意の妥当性について検証していく態度である。トップダウンの思考実験を、ボトムアップの研究の蓄積によって打破していく研究姿勢は、使用基盤モデルの理念そのものである。

参考文献

Dąbrowska, Ewa and Divjak, Dagmar (2019) *Cognitive Linguistics: Foundations of Language*. De Gruyter Mouton: Berlin.
/ Lakoff, George (1990) "The Invariance Hypothesis: is abstract reason based on image-schemas?" *Cognitive Linguistics*, 1(1): 39-74. / 永田亮・大谷直輝・高村大也・川崎義史 (2022) 「言語処理的アプローチによる *better off* 構文の定着過程の説明」言語処理学会第 28 回年次大会. / 大谷直輝 (2022) 「*better off* 構文の定着過程に関する認知言語学的考察」言語処理学会第 28 回年次大会. / Searle, John (1980) "Minds, Brains and Programs," *Behavioral and Brain Sciences*, 3 (3): 417-457.

ⁱ 一例を挙げると、単語を置き換えた高次元のベクトルが、何らかの認知作用や言語学が想定するメタ言語に対応するなど、自然言語処理の分析の過程を見ていくことで、その一部に対応するような人間の認知的な傾向が明らかになる可能性が考えられる。

ⁱⁱ 大谷 (2022) が提唱した認知言語学的な仮説の検証に深層学習を用いた、永田・大谷・高村・川崎 (2022) では、仮説の検証のみならず、データ構築における効率化、仮説の修正、新たな仮説の構築などに深層構造の手法が役立つ点が示された。