

深層学習を用いた構文文法の実証的な研究の可能性を探る

——better off 構文を例にして——

大谷 直輝 永田 亮 高村 大也 川崎 義史
東京外国語大学 甲南大学 産業技術総合研究所 東京大学

【要旨】 本稿では、*better off* 構文と呼ばれる不規則性を含むイディオム構文がどのような過程を経て現代英語に定着したかを認知言語学・コーパス言語学・自然言語処理の学際的な観点から検証する。同時に、*better off* 構文の分析を通じて、近年著しく発展した深層学習を用いた構文文法の実証的な研究手法の提示を試みる。本稿では、通時的コーパスを用いた *better off* 構文の成り立ちに関する定量的な調査から得られた結果を認知言語学の観点から考察することで、*better off* が字義的な用法から四つの段階を経て構文化したとする仮説を提示し、その後、その仮説を実証するために人手の分類に加え、*better off* が用いられる文脈情報から得られる単語ベクトルを用いた検証を行う。さらに、分析全体を通じて、1) 自然言語処理の技術を用いた実証的な言語の研究サイクル、2) 用法基盤モデルと深層学習に基づく言語モデルの類似点、3) 学際的な収束証拠の一つとして深層学習を用いる可能性についても検討を行う*。

キーワード： *better off* 構文、構文文法、用法基盤モデル、深層学習、単語ベクトル

1. はじめに

本稿の目的は以下の三点である。一点目は、*better off* 構文と呼ばれる不規則性を含むイディオム構文がどのような過程を経て英語に定着したかを認知言語学・コーパス言語学・自然言語処理の学際的な観点から検証することである。この目的を達成するため、通時的なコーパスを用いた *better off* の調査から得られた結果を認知言語学の観点から考察することで仮説を生成し、その仮説を言語学の観点から検証する。さらに、自然言語処理と深層学習を利用したデータ駆動の分析を用いて検証結果を補強することを目指す。同時に、*better off* 構文の研究を通じて、二点目として、分析の手順を、1) 現象の記述、2) コーパス調査、3) 仮説の生成、4) 仮説の検証という四段階に分けて論じることで、言語を実証的に分析するための研究サイクルを提示する。さらに、三点目として、実証的な言語分析をする上で2010年代から

* 本稿は *better off* 構文に関する著者の一連の通時的・共時的研究をまとめ、実証的な言語研究手法を提示するという観点から大幅に加筆・修正したものである。学会や研究会等で多くの研究者から貴重なコメントをいただいた。査読を通じて、匿名の2名の査読者から多くのご助言をいただいた。また、佐近優太氏、稲生秀俊氏には、原稿全体に目を通していただきコメントをいただいた。なお本研究は科研費（基盤研究（C）：19K00657）の助成を受けて行われている。

急速に発展した深層学習に基づく自然言語処理のアプローチが利用できる可能性について検討する。

現代英語には、(1)のように、形容詞 *better off* の叙述用法に非時間的關係 (atemporal relation) (Langacker 1987: Ch. 6) を表す様々な要素が後続し、全体が「～したほうがましだ」という意味を表す用法が広範に観察される。

- (1) (a) She is *better off* *without me*. (彼女は私といない方がましである)
 (b) I'd be *better off* *at home with all those kids!* (私は家でこの子たちと居た方がましだろう)
 (c) Maybe I'd be *better off* *in jail*. (おそらく、私は刑務所に居た方がましだろう)
 (d) He is *better off* *buying it*. (彼はそれを買った方がましだ)
 (e) I'd be *better off* *left alone*. (私は一人残された方がましだろう)
 (f) I'd be *better off* *{dead/alone}*. (私は {死んだ/一人の} 方がましだろう)

better off の語源は「豊かな」という意味を表す *well off* の比較級であるが、(1) では、*better off* の後に、様々な品詞を主要部とする非時間的關係を表す要素(下線部を参照)が続いている。例えば、(1a) (1b) (1c) には前置詞句が、(1d) と (1e) には現在分詞と過去分詞が、(1f) には形容詞が、*better off* の後に現れている。これらの要素は、意味の観点から見ると、主語の現在の状態とは異なる仮想的な状態を表すという共通点が見られる。例えば、(1a) では、「彼女は現時点では私と一緒にいるが、その状態に比べて、私とは一緒にいない(仮想的な)状態の方がましである」ことを表している。

本稿の構成は以下の通りである。2 節では、本稿の理論的な背景となる構文文法の主張や変遷を概観しながら、近年、用法基盤モデル(または、使用基盤(依拠)モデル) (the usage-based model) に基づく構文研究が盛んになってきた背景や、現代の構文研究に見られる課題を記す。3 節では、2010 年以降に飛躍的に進化した深層学習の特性を概観すると同時に、深層学習やそこから得られる単語ベクトル(分散表現)が認知言語学の観点から見るとどのような意味を持つかを検討する。4 節では、古英語や中英語には見られなかった *better off* 構文(つまり、*better off* に主語の仮想的な状態を表す要素が後続する用法)が、語源となる形容詞 *well-off* の比較級から、どのような経路を経て英語に出現したかを明らかにすることを試みる。4.1 節では、*better off* 構文の成り立ちを 4 つの段階に分けて検証をするという本稿の方法論を示す。4.2 節(第一段階)では、*better off* 構文に見られる構文的な特徴を記述する。4.3 節(第二段階)では、英語の歴史コーパスである *Corpus of Historical American English* (以下、COHA) における *better off* が共起する要素の変化等を考察することで、*better off* 構文の成り立ちに関する定量的な調査を行う。4.4 節(第三段階)では、調査結果を認知言語学の観点から論じることで、*better off* 構文は字義的な用法から 4 つのステージを経て定着したとする仮説を提示する。4.5 節(第四段階)では、人手による *better off* の用例の分類と、深層学習に基づいた言語モデルの一種

である BERT による *better off* の用例のクラスタリング¹との比較を通して、仮説の妥当性を検討することを試みる。4.6 節では、4 節の議論をまとめ、実証的な言語分析サイクルを提示すると同時に、言語分析サイクルを構築するうえで、自然言語処理がどのように貢献するかを検討する。5 節では、本稿の議論をまとめた後、深層学習時代の構文研究という観点から、1) 深層学習を言語分析に用いる注意点と、2) 構文文法や認知言語学が自然言語処理や深層学習の分野へ貢献できる可能性を検討する。

2. 構文文法の変遷：形式と意味の対から用法基盤へ

2 節では、本稿の背景となる構文文法の研究の動向をおもに用法基盤モデルの観点から振り返る。*let alone* 構文の特性を論じた Fillmore, Kay and O'Connor (1988) を構文文法の始まりと考えると、現在までに構文文法が誕生してすでに 35 年の月日が流れている。構文文法は、形式と意味の対からなる言語記号 (linguistic sign) というソシュールが提唱した概念を語だけではなく句や節のレベルにまで広げた「構文」という概念を提示し、言語知識は形式と意味の対からなる構文のネットワークからなるとする主張を行った。この主張は、当時の通説となっていた「言語知識はおもに語彙と文法からなり、その例外がイディオムである」という言語観に対して、「語彙と文法は言語知識という尺度の両端にあるものであり、両者の間に横たわる膨大に存在する (様々な抽象性や複雑性を持つ) イディオムこそが人間の言語知識の中心である」と主張する点で革命的であった。さらに、構文文法は、Goldberg (1995) を境に、項構造構文という中心的な研究テーマを見つけ、動詞が表す事態構造からは推測が難しいため、項構造構文によって認可されると考えられる、動詞の創造的使用の分析が広がっていった (Boas 2003; 小野 2007; Iwata 2008; Kim and Michaelis 2020)²。

現代の構文文法の研究範囲は、初期の構文文法の研究範囲に比べ、少なくとも以下の三点で広がりを見せている。一点目は、構文を構成する 2 つの要素である形式と意味が表す範囲である。現代の構文文法が扱う形式には、音声や文字のような伝統的な言語的な形式だけでなく、イントネーションユニットをはじめとしたパラ言語的な要素 (Cruttenden 1997)、さらには、近年研究が多くなされているマルチモーダル構文 (Steen and Turner 2013) の分析対象であるジェスチャー、指差し、表情、視線等も含まれる。一方、意味についても、意味論的な意味だけでなく、語用論的な機能や、対人関係や言語使用域をはじめとした文脈的な意味も含まれる (Verhagen

¹ クラスタリングとはデータを特定の規則に基づいていくつかのクラスター (グループ) に分類することを指す。

² 例えば、*He sneezed the napkin off the table.* や *He cooked me the meal.* 等の文は、自動詞の *sneeze* (くしゃみをする) や他動詞の *cook* (料理をする) という動詞が持つ意味より、文全体が複雑な事態を表しているが、これらは使役移動構文や二重目的語構文のような項構造構文によって認可されると考える。

2007)。二点目は、構文の長さや複雑性である。伝統的に、構文の研究は、語、句、文レベルの構文の分析が中心であった。一方で、コーパスを用いた文脈調査の方法の確立等の影響によって、文を超えた談話レベルで機能する構文や、隣接発話対等の談話レベルの抽象的な構文も研究対象となっている (Tomasello 1998, 2003a; 鈴木・秦・横森 2017; 中山・大谷 2020)。同時に、語よりも小さな単位として、屈折や派生を扱う形態論へも拡大している (Booij 2010)。三点目は、構文の抽象性である。初期の頃は、人間の言語知識に相当すると仮定される構文のネットワーク (すなわち、Construct-icon) の中でも、二重目的語構文 (SVOO) のようなすべてがオープンスロットである抽象的な項構造構文と、*let alone* 構文のような具体性の高いイディオム構文が研究の中心であったが、近年は、両者の間を埋めるような、一部にオープンスロットを含む中間的な構文 (meso-constructions) の研究が多くなっている (Boas 2003; Stefanowitsch and Gries 2003)。同時に無意味構文 (meaningless construction) と呼ばれる、対応する意味を見つけないのが難しい抽象性が極めて高い構文 (句構造、主語述語構文など) についても検討がなされている (Hilpert 2019)。

このような研究の拡大に大きな影響を与えているのが用法基盤モデル、すなわち、「言語知識は言語使用の中で繰り返し現れるパターンから出現する」という考え方である (Bybee 1985; Langacker 1988; Tomasello 2003b)。初期の構文文法には必ずしもこの傾向は見られなかったが、近年は、言語使用からボトムアップに立ち現れる構文の特性を探るという研究の姿勢が一般的に共有されている³。このことは、Goldberg (2006) で提示された、現時点で最も一般的な構文の定義において、従来から主張されている「構文は形式と意味の対からなる単位」という定義に、「頻出するパターンはそれ自体が構文として蓄えられる」という一文が加えられたことから分かる。

一方で、構文文法が現在直面している理論的な課題もある。ここでは、本論文との関連で2つの課題を述べる。一点目は、Lakoff (1990) で提唱された認知言語学の分析における前提となるべき、1) 言語のあらゆる側面に対して適用できるような、一般性の高い原理を提案すべきであるという「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)」が理念にとどまっている点が挙げられる。たとえば、構文文法では、言語知識を人間の脳に記憶された心的なものとして捉えているにも関わらず、分析の多くが言語学という分野内にとどまり、心理学や脳科学といった隣接分野の研究成果を取り入れる、あるいは、隣接分野の方法論を用いた検証を行うなどの学際的な研究が進んでいない現状にあ

³ 参考として挙げると、近年出版された構文文法の教科書、Hilpert (2019)、大谷 (2019)、坪井・早瀬 (2020)、Hoffmann (2022)、Croft (2022) はほとんどが用法基盤モデルに基づいている。また、例外として記号基盤構文文法 (Sign-based Construction Grammar: SBCxG) の観点から書かれた、Kim and Michaelis (2020) がある。

る。構文文法が認知科学の一端を担い、人間の心の働きの一部である言語知識のありようの解明を目指す学問である以上、その論証には、言語における傾向の提示を超えた、学際的な収束証拠 (converging evidence) を提示することが求められる (Schönefeld 2011)。

二点目は、理論の実装という問題である。構文文法では、量的転回などを受け、コーパスや心理実験等の定量化できる手法を用いた言語分析が広がっており、身体的構文文法 (Embodied Construction Grammar (ECG)) や流動的構文文法 (Fluid Construction Grammar (FCG)) など一部に自然言語処理や計算言語学との接近を試みる学派も存在する (Bergen and Chang 2005; Steels 2013)。一方で、ニューラルネットを利用して構文⁴を計算言語学的にモデル化し、実装をするという試みはまだ広がっているとは言えない⁵。

3. 自然言語処理の認知言語学的解釈

本節では、本稿で学際的な研究を行う自然言語処理における近年のパラダイムの変遷や必要不可欠な要素技術となりつつある単語ベクトルを認知言語学の観点から見るとどのように捉えられるかについて検討する。特に、事前に規則を与えるのではなく、与えられたデータのみから、所望の出力を得るための学習を行うという深層学習の枠組みが、「言語知識は言語使用の中で繰り返し現れるパターンから出現する」という用法基盤モデルの考え方に通じる部分があり親和性が高いことを議論する。

3.1. 自然言語処理のパラダイム変遷と用法基盤モデル

自然言語処理では、与えられた入力に対して何らかの処理を施して所望の出力を得る。例えば、自動翻訳 (自然言語処理では機械翻訳) では、与えられた、ある言語の文を別の言語に翻訳して出力する。所望の出力をどのような処理により得るかという観点から自然言語処理のパラダイム変遷を追っていくと、3つの段階に分けることができる⁶。最初は、2000年以前に行われていた、人が作成した解析の規則を用いて処理を行うという段階である。例えば、機械翻訳の場合、品詞、統語、意味などを解析する規則を機械に与え翻訳を行う。そのような規則は、言語学の知見や、内省 (reflection) のような言語学的方法論を用いて作成することが主流であった。したがって、この段階の自然言語処理は言語学への依存度が高かったと言える。

⁴ ここでの「構文」は、自然言語処理で用いられる構文はなく、構文文法が規定する形式と意味の対としての構文のことである。

⁵ 第16回国際認知言語学会 (ICLC16) (2023年8月7日から11日@デュッセルドルフ) においても、深層学習や大規模言語モデルを用いた言語研究の可能性に触れた研究が数件あった程度であった (Hasebe 2023; Torrent, Almeida and Turner 2023, など)。

⁶ ただし、ここでの3つの段階は、大まかなものであり、各段階の方法が別の段階でも使われることがある。三段階目の深層学習ベースの自然言語処理でも、依然、言語学の知見が重要な役割を果たすこともある。

次に、2000年代以降の、データ駆動の処理が盛んになった段階である。ここでは、処理のための規則を直接作成するのではなく、コーパスに言語学的な情報を付与し、そこから機械学習を用いて規則を得る。例えば、品詞解析であれば、コーパスに品詞情報を付与する。そのコーパスに機械学習を適用して品詞解析のための規則を自動生成する⁷。ここで重要なことは、規則は自動的に得られるようになったものの、品詞、句、統語などの言語的な知識の存在が前提とされていることである（コーパスに対する情報付与は、言語学的な知識に基づいて行われる）。したがって、前段階に比べると低くなったものの、この段階でも依然、言語学的な知識に依存していると言える。2010年代に入り深層学習が導入されると、品詞解析、構文解析などの途中の解析を介さず、入力から直接所望の出力を得るアプローチが主流となった⁸。例えば、上述の機械翻訳の例では、翻訳対象文を入力すると訳文のみ出力される。一方、以前の段階では、どの語を主語と認識しているかを確認できるなど、必要に応じて、品詞解析や統語解析の結果を出力することが可能であった。したがって、明示的な解析のための規則は存在せず、規則は、ニューラルネット⁹内の状態として抽象的に保持されている。以上のように、この段階になると、品詞や統語というものを仮定する必要はなくなり、言語学的な知識への依存が低下したと言える。

この三つの段階を、認知言語学の観点から考察すると、以下のような2つの見方が可能である。一点目は、自然言語処理に対する言語学の影響力の低下である。言語学者の直観が作り上げた品詞や構文を用いることで、自然言語を処理することが可能となった第一期に対して、第二期では、品詞や構文等の情報付きのコーパスからのデータを用いて、機械が解析ルールを作り上げていった。第三期になると、言語学者が直観によって規定をする言語学的な情報が付与されていない状態であっても、機械がいわゆる「文法」に対応するような言語の生成や解析のための規則を生言語データから学習するようになった。

二点目は、深層学習に基づく自然言語処理のアプローチ（すなわち、事前に言語学的な知識を与えずに¹⁰、使用例（データ）からのみ、処理方法を学習するというアプローチ）と、用法基盤的な考え方には共通点がある点である。第1に、大量の言語データが必要になるものの、品詞、文法関係、統語規則等を事前にデータに付与しなくても、生の言語データからそれらを学習できることは、言語使用の中からそれらを学べると主張する用法基盤モデルの考え方に似ていると言える。第2に、

⁷ 得られた規則は必ずしも人間が理解可能な形になっていないこともある。

⁸ それ以前にも、扱った問題によっては、途中の解析を行わないアプローチがとられることはあったが、深層学習導入後は、幅広い問題についてこのアプローチが主流となった。

⁹ 流派にもよるが、深層学習とニューラルネット（もしくは、ニューラルネットワーク）は同じような意味で用いられる。本稿では、深層学習は概念全体、ニューラルネットは、深層学習により実現された手法とプログラムのことを指す。

¹⁰ 厳密には、広い意味での言語学的な知識を暗に仮定することはある。例えば、機械翻訳の場合、一文ずつ翻訳することが多い。これは文という言語学的な単位を暗に仮定していると言える。

文法も語彙もイディオムも生の言語データから同様のニューラルネットを用いて学習するということは、文法は生得的に獲得し¹¹、語彙は学習で習得をするという異なる習得経路の違いを仮定せずとも、文法も語彙も言語使用の中から一般的な認知能力を用いて習得していくと考える用法基盤モデルと類似した考え方とみなせる。第3に、深層学習では、生の入力データから言語を解析する規則を作り上げる際に、文脈の要素を均等に扱うわけではなく、文脈の中で重要な部分とそうでない部分を区別し、重要な部分に特に注目をするような設計がなされている (Vaswani et al. 2017)¹² 点も、発話の中で重要な部分とそうでない部分を分けながら認識する人間の認知機能に近いと言える。

現代の深層学習に基づく自然言語処理の考え方を用法基盤モデルと比較すると、以下のような類似点が見られる。

表1 深層学習に基づく自然言語処理と用法基盤モデルの類似点

- | |
|---|
| <ol style="list-style-type: none"> 1. 発話された文 (uttered sentences) が重要なインプットになる。 2. 規則や言語知識は前もって与えられず、使用の中から学習することができる。 3. 言語使用から繰り返し用いられるパターンが抽出されて言語知識となる。 4. 文脈を均等ではなく、重要な部分とそうでない部分を区別しながら認識する。 |
|---|

3.2. 単語ベクトルと言語知識のありよう

現代の自然言語処理におけるブレイクスルーをもう一つ挙げると、単語をはじめとした言語ユニットの意味の扱いがある。従来は、単語等は記号として処理されていたが、現在は、単語が持つ属性をより柔軟にあつかうために、単語を数値列 (単語ベクトル) に置き換えて処理する。単語ベクトル (分散表現とも呼ばれる) は、深層学習時代の自然言語処理における必要不可欠な要素技術となっており、代表的なものに、Word2Vec, BERT などがある¹³。単語を記号とみなすのではなく数値列に置き換えることによって、類義性や統語的属性のような単語間の関係を定量的に捉えることができる。また、ベクトルは、語だけではなく、句、節、文、パラグラフ、文書といったより大きな構造に対しても得ることができる。

このような語や文の特性を数値列として扱うベクトルは「言語知識は体系化された構文のネットワーク」と考える用法基盤モデルの考え方と相性がいいように思われる。ベクトルは、ある空間上の一点として解釈することができる。したがって、単語集合や文集合に対するベクトルは空間上に配置された点の集合となる。また、

¹¹ 自然言語処理の観点からも、事前知識を与えずとも、入力から所望の出力を得るシステムを実現できるという点で、生得的なものを仮定せず、使用 (データ) から全てを学習できるとみることでもできる。一方で、ある特定の能力を有するシステムを実現するためには、ある特定の構造を持ったニューラルネットが必要となることも明らかになりつつある。これは生得的なものを仮定していると解釈することもできる。この点については更なる研究が待たれる。

¹² アテンションと呼ばれ、人間の認知的な注意を模倣するように設計されたものである。

¹³ Word2Vec については Mikolov, Chen, Corrado and Dean (2013) を、BERT については Devlin, Chang, Lee and Toutanova (2019) を参照。

上述のように、単語ベクトル間の類似度を計算することが可能であり、類似した単語（すなわち点）を結ぶとある種のネットワークと解釈することもできる。単語ベクトルを用いることで、類義性などの単語間の関係を捉えたり、語・句・節が表す意味をベクトルという同様の観点から捉えたりすることが可能となる。また、言語ユニットを置き換えた数値列では、言語学における語彙的な意味と文脈的な意味を区別せず、両者を文脈的な情報とみなす点も、両者を区別せず連続的なものとみなす用法基盤モデルの言語観と合致しているように見える。

3.3. 次節に向けて

自然言語処理の分野では、現在、深層学習を用いて様々な対象が分析されているものの、深層学習の結果生成される数値列（単語ベクトル）が実際に何を意味しているかについては言語処理学者の中でもよく分かっていないという実情がある。また、Searle (1980) が指摘するように、人工知能が人間の脳を模倣しているかを検証することは不可能であるとの意見も存在する¹⁴。そのような立場に立つと、深層学習を人間の言語知識に見立てて、言語研究に用いるという本稿の前提は危ういように感じられるかもしれない。一方で、本節で見てきたように、以前の自然言語処理と現代の深層学習を比較すると、表1が示すように、深層学習と用法基盤モデルの考え方には多くの類似点があるように見える。そこで、本稿では、自然言語処理の分野に見られる様々な懸念について十分に留意しつつも¹⁵、深層学習を用いた言語研究の姿を4.5節で検討する。

4. *better off* 構文の成り立ちに関する実証的な研究

4.1. 研究の方法論

4節では、*better off* 構文の成り立ちを論じる実証的な研究を、4つの段階に分けて提示する。第一段階（4.2節）では、*better off* 構文という不規則性を内在する構文が持つ特徴を形式と意味の観点から記述する。第二段階（4.3節）では、*COHA* を用いて *better off* 構文がどのような過程を経て誕生したかを調査する。第三段階（4.4節）では、コーパスから得られた調査結果を認知言語学の観点から解釈することで、*better off* 構文は語源となる *well off*（裕福な）の比較級から4つのステージを経て英語に定着したという仮説を示す。第四段階（4.5節）では、第三段階で提示した仮説を、人手による *better off* の用例の分類と、深層学習を用いた *better off* の用例のクラスタリングに基づくグループ化の比較を通して検証することを試みる。最後に、4.6節では、実証的な言語分析のサイクルを提示して、その有用性を示す。

本節の方法論を表にまとめると以下のようなようになる。

¹⁴ 一方、脳の働きを深層学習と対応付けする研究は近年着実に進められている（Horikawa and Kamitani 2022, など）。

¹⁵ 深層学習と用法基盤モデルに見られる相違点や、深層学習を人間の言語知識に見立てた分析の注意点や危険性については、5.1節で論じる。

表2 本稿における *better off* の分析の段階

第一段階	<i>better off</i> 構文の特徴の記述
Hilpert (2019) が提唱する構文を認定する4つの基準に基づいて <i>better off</i> 構文が持つ形式的・意味的特性を記述する。	
第二段階	<i>better off</i> 構文の成り立ちに関するコーパス調査
歴史コーパスの COHA から網羅的に収集した例に変数を付与することで <i>better off</i> の生起環境が1820年以降、アメリカ英語でどのように変化したかを量的に捉える。	
第三段階	調査結果に基づいた <i>better off</i> 構文の成り立ちに関する仮説の生成
コーパス調査の結果を認知言語学の観点から論じることで、字義的な用法から4つの段階を経て <i>better off</i> 構文が慣習化したという仮説を提示する。	
第四段階	仮説の検証と深層学習を用いた検証の補強
深層学習の BERT を用いて各用例を単語ベクトルに変換したものをクラスタリングした結果を、コーパス調査の結果と比較することで、仮説の実証を試みる。	

4.2. 第一段階：*better off* 構文の特徴¹⁶

better off 構文が歴史的にどのような過程を経て英語に定着したかを検証するための第一歩として、まず、*better off* 構文に見られる特異な性質を示す必要がある。そこで、第一段階では、構文文法の観点から、*better off* 構文に見られる構文的な特性を明らかにする。Hilpert (2019) では、構文を認定する基準として、1) 形式面の不規則性、2) 意味面の非構成性、3) 表現独自の制約、4) コロケーションの選好性を挙げている。本稿では、この基準に従って、*better off* 構文が持つ特性を示していく。以下、再掲した (1) に基づいて議論を進める。

- (1) (a) She is better off *without me*.
- (b) I'd be better off *at home with all those kids!*
- (c) Maybe I'd be better off *in jail*.
- (d) He is better off *buying it*.
- (e) I'd be better off *left alone*.
- (f) I'd be better off *{dead/alone}*.

最初に、*better off* 構文の形式的な特徴に注目する。*better off* 構文は、形容詞 *better off* の叙述用法に、非時間的關係が主要部となる様々な要素（前置詞句、現在分詞、過去分詞、to 不定詞、形容詞など）(XP と呼ぶ) が後続するという特徴がある。

- (2) *better off* 構文の形式的特徴
NP_[subject] + *be* + *better off* + XP

(2) が示す *better off* 構文の形式的な特徴には、2つの点で標準的な文法からは逸脱した特性が見られる。一点目は、形容詞 *better off* の後に補語句とも解釈されるよ

¹⁶ 本節の内容は大谷 (2018) に基づいている。

うな標準的な文法では分類できない要素が後続する点である。この要素は文の付加句ではなく、*better off*の補語句に近いものとみなせる。付加句であるならばその要素が主節の前に現れることもできるはずであるが、主節の前にXPが現れた場合、*better off*構文が持っている「仮想的な状態の方がまだ」という意味は生じない。すなわち、XPが*better off*直後に出てきた場合のみに非構成的な意味が見られるため、XPは主節に対する付加句ではなく、*better off*句内の要素とみなした方が適切と言える。例えば、(1a)の前置詞句が、主節の前に置かれた場合、「私がいなければ、彼女は調子が良い」というような意味になり、「いない方がまだ」というニュアンスは生じない。さらに、(1a) (1b) (1c)の前置詞句については文の付加句としても解釈が可能であるが、(1d) (1e) (1f)の分詞や形容詞は通常は付加句とはならない要素である。そのため、(1f)の形容詞が主節の前に現れると文全体が非文になり、(1d) (1e)の分詞節が主節の前に現れる場合、分詞節として解釈され、分詞節と主節の間に一般的にカンマが現れることとなる。以上のことから本稿では、*better off*に後続する要素(XP)を*better off*の補語句と呼ぶ¹⁷。二点目は、XPの主要部となる品詞(前置詞、分詞、形容詞)を統括するカテゴリーが伝統的な品詞論においては認識されていない点である。対照的に、認知文法においてこの統括的なカテゴリーは、動詞が表す時間的な関係(temporal relations)と対立する非時間的な関係を表すカテゴリーとして認識されている¹⁸。

次に、意味的な面に注目を見ると、XPは、主語が現在おかれた惨めな状況に比べてまだましと考えられる主語の仮想的な状態を表し、*better off*構文全体では「主語はXPの状態の方がまだ」という非構成的な意味を表す。*better off* XPで「XPが表す状態の方がまだ」という意味を表すため、(be) *better off* XPは全体で、根源的モダリティ(root modality)を表す助動詞句と類似した機能を持つ。また、語用論的な意味に注目すると、XPが表す事態は必ずしも一般的に好ましいわけではないため、誇張(*I'm better off dead*)や皮肉(*I'd be better off in jail.*)が生じやすくなる。さらに、使用される言語使用域に注目すると、話し言葉や砕けた会話で用いられやすいという特徴が見られる。

また、Hilpert (2019)で挙げられた3つ目と4つ目の構文と認定する基準に注目を見ると、*better off*が補語句を取る場合(すなわち*better off*構文)と取らない場合に顕著な差が見られる。まず、3)の「表現に独自の制約がある」という特徴については、*better off*に補語句が後続する場合、叙述用法のみで現れ、少数の例外を除

¹⁷ ただし、形容詞に補語句が続く場合、*I'm fond of sweets.*のように一般的には前置詞をとまなうため、XPを補語句とみなした場合でも、不規則性が含まれることとなりXPは典型的な補語句とは言えない。

¹⁸ 査読者が指摘するように、この統括的なカテゴリーは伝統的な品詞論では自然類を成さないが、時間関係を取り入れた認知文法では自然類として考慮できるため、認知文法に基づく品詞論の妥当性の根拠ともなる。

いて限定用法では使わないという特徴が見られる¹⁹。一方、形容詞 *better off* が補語句を取らない場合、叙述用法 (*He is now better off*)、限定用法 (*the better off person*)、名詞的用法 (*the better off*) が確認される。

また、4) 「 कोरोケーションにおける選好性がある」という特徴については、*better off* が補語句を取る場合は、補語句をとらない場合に比べて、一人称や二人称との共起や、助動詞との共起が有意に高いことが分かっている。つまり、*better off* 構文には、主語_[1st person/2nd person] (auxiliary) *be better off* XP のような頻出のパターンがあることが分かる。このような表現は、話し言葉で多用され、*You'd be better off going home* や *We'd be better off without him.* のように、提案や勧誘等を表す発話行為として機能する用法である。

4.3. 第二段階：*better off* 構文の成り立ち²⁰

第二段階では、*better off* 構文が持つ特性がどのような過程を経て *well off* (裕福な) の比較級から出現したかの調査を行う。最初に、第一段階の調査を基にして、*better off* における *better off* 構文と叙述用法が持つ特性の違いをまとめる。形式面では、*better off* が補語句を取らない叙述用法の場合、*better off* 自体が文の補語として主語の状態を表すが (例 *I'm now better off than before*)、*better off* 構文では、*better off* ではなく、*better off* に後続する補語句が主語の状態を表す (*I'm better off dead*)。また、叙述用法の *better off* は、主語が比較対象より「豊かな」あるいは「より良い」状態を表すが、*better off* 構文では、*better off* の補語句が主語の仮想的な状態を指すため、*better off* には「(その状態の方が) ました」という意味を表す、助動詞の根源的モダリティを表す用法と類似した機能があると言える²¹。

19世紀なかばの典型的な *better off* の使用は以下のような叙述用法であり、*better off* は、修飾する要素が基準となる時点に比べて裕福であることを表していた。

- (3) (a) I was better off once, and then I had plenty of friends. [sic] [1848]
 (b) There! you are no better off than you were before! [1865]

一方、COHAを確認したところ、19世紀の *better off* には補語句を取る用法はほとんど観察されなかった。そこで、本稿では、*better off* に生じた様々な構文的な変化の違いを捉えるため、COHAを利用して19世紀以降、*better off* に生じた使用文脈の変化を通時的に調査する。

COHAは1810年代から2010年代のアメリカ英語約4.7億語からなる通時的なコーパスである。各文書には作成年とジャンルの情報が付与されている。本研究では、

¹⁹ COHAでは、*better off* 構文の限定用法での使用は見つからなかったが、*the better-off-dead person* のような、ハイフンをともなった例は少数ながらインターネットで観察された。

²⁰ 本節の内容は Otani (2021) に基づいている。

²¹ 根源的モダリティは「義務」「許可」「可能」など「そうあるべき」世界に関わるため、「ましな」世界を表す *better off* も根源的モダリティに関わると言える。

COHAから *better off* の全例を抽出し、それらの例に対して、表3の変数をコーディングし、変数間の関係を定量的に処理することで、*better off* が使われる文脈の変化を明らかにする。

表3 *better off* にコーディングした変数と値

変数	値
文法役割	(1) 叙述用法, (2) 限定用法, (3) 名詞用法
補語句	(1) 補語句有り, (2) 補語句無し
主語の人称	(1) 一人称, (2) 二人称, (3) 三人称
助動詞	would, will, should, can, など
than	(1) 有り, (2) 無し
if節	(1) 有り, (2) 無し
意味	(1) 字義的意味, (2) 非字義的意味
ジャンル	(1) Fiction, (2) Popular Magazines, (3) Newspapers, (4) Non-Fiction Books
出現年	1810–2009

COHAには *better off* が全部で2642例見られ、その中で、*better off* 構文の基盤となる叙述用法は2527例見られた。さらに、*better off* に後続する要素の中で、その要素が文の付加句ではなく、確実に *better off* の補語句として解釈できる、補語句の主要部が分詞句と形容詞である例は354例であった。*better off* と共起する要素の通時的な変化を図1で示す。

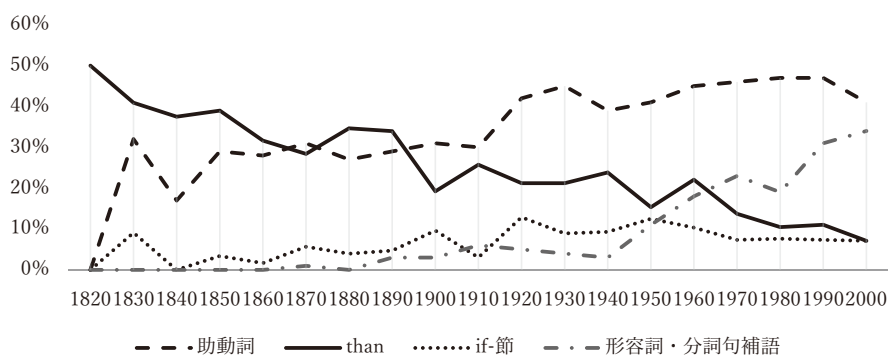


図1 *better off* に共起する要素の変化

better off の通時的な調査の結果、*better off* の使用文脈に現れた変化には以下の傾向が見られた。第1に、*better off* と *than* との共起頻度は一貫して減少していた。これは、当初は *well off* の比較級という語源的な意味を強く保持していた *better off* に「比較」の意味が薄れてきたことを意味する。第2に、*better off* と認識モダリティを表す助動詞との共起頻度は、1920年以降、増化した。これは、*better off* を含む文にお

いて、仮想的な事態を表す文の割合が増えてきたことを意味する。第3に、*better off*に後続するXPの位置に、補語句として解釈せざるを得ない形容詞と分詞句が1950年以降、増え始めた。いわゆる、標準的な文法からは逸脱しているこのような要素が増えてきていることは、*better off*構文の定着度が高まってきたことを意味する。第4に、字義的な意味の割合に注目すると、19世紀の*better off*は多くが字義的な「より裕福な」という意味（例 *He is financially better off than her*）を表していたが、2000年代に入ると、字義的な意味は17%まで減少していた。これは、字義的な意味がかなり薄れてきていることを意味する。以上のように、*better off*の使用文脈は19世紀から現在まで徐々に変化を続けており、変化が積もった結果、19世紀にはほとんど見られなかった*better off*が補語句を取る新しい用法が現代英語に定着していることが分かる。

次節では、コーパス調査の結果を認知言語学の観点から論じることで、*better off*構文がどのように誕生したかを論じる仮説を提示する。特に、*better off*の補語句という他に例を見ない現象がどのように誕生したかに注目をする。

4.4. 第三段階：仮説の生成²²

第三段階では、第二段階でのコーパス調査の結果に基づいて、「*better off*構文は、4つのステージを経る中で、*better off*に後続する文の付加句が補語句として再分析(analysis)されることで誕生した」と主張する仮説を提唱する。具体的には、表4が示すように、字義的な意味を表していた*better off*に、(i) 意味の一般化、(ii) 仮想状況(=セッティング²³)の前景化、(iii) セッティングの補語句化、が起こることで、他に類例がない、主語の仮想的な状態を表す補語句を持つ*better off*構文が誕生したと考える。以下、4つのステージ別に、*better off*構文が慣習化するまでの過程を見ていく。

表4 *better off*構文が慣習化するまでの4ステージ²⁴

第一ステージ：意味「より裕福な」(字義的)
↓ 意味の一般化
第二ステージ：意味「より良い」
↓ 仮想的な状況の前景化・比較対象の背景化
第三ステージ：意味「特定の状況でより良い」
↓ 仮想的な状況を表す要素の補語句化
第四ステージ：意味「～の方がました」(構文誕生)

²² 本節の内容は大谷(2022)に基づいている。

²³ セッティングとは、事態が生じる状況と定義される(Langacker 2008)。

²⁴ 表4は、*better off*が4つのステージを順に移行したわけではなく、*better off*が各ステージを経る中で、そのステージに対応する意味や機能を加えていき、最終的に*better off*に複数の意味や機能が層(layering)をなして同時に存在することを表す(Hopper 1991)。

第一ステージは, *better off*が *well off*の比較級として字義的な「より裕福な」という意味を表していた段階である。主語は比較対象より, 経済的に裕福な状態にある。

- (4) (a) They are better off here than on most plantations. [1868]
 (b) … the entire community will be one third better off than before. [1870]
 (c) I'm better off than you are, and I can spare the clothes … [1868]

(4a) では, 同一の集団の二つの状態が, (4b) では, 同一の共同体の二つの状態が, (4c) では二人の人物の状態が比較され, 主語が *than* 句によって明示的に示すことができる比較対象より裕福な状態であることを表す。この用法では, *better off*は字義的な「より裕福な」という意味を表す。また, 比較の意味が強く残っていることは *than* との共起からも示唆される。

第二ステージは「より裕福な」から「より良い」へ, 意味の一般化が起こる段階である。

- (5) (a) I'm better off at home. [1920]
 (b) Clearly I am better off in my childless condition, and … [1899]
 (c) I think your baby is a good deal better off without you. [1913]

(5) の各例では, 二つの状態が比べられ, 主語の状態が比較対象に比べて「より良い」ことが示される。ここでは, *better off*に後続する *at home, in my childless condition, without you* のような前置詞句が, 主語がより良い状態となる背景的な状況を表している。例えば, (5a) では「家に居る」, (5b) では「子供がいない」, (5c) では「あなたがいない」という状況において, 主語は「より良い」状態となる。この前置詞句が表す状況は, 事態が生じるセッティングとしてみなすことができる。セッティングは通常, 主節に対する付加句として現れるものである。(5a) (5b) (5c) の各文においても「家に居ない」「子供がいる」「あなたがいる」という比較対象は存在するが, 背景化され言語化はされていない。

第三ステージは, 主語が置かれた仮想的な状況 (セッティング) が前景化される段階である。第二ステージでは, *better off*に意味の一般化が起こり「より良い」という意味を表すようになったが, 「より良い」状況を表すには, 2つの状況が比較される必要がある。第三ステージになると, この二つの状況のうち前置詞句や *if* 節のような付加句的な要素によって表される状況が現実とは異なる仮想状況を表す, (6) のような用法が増え始める²⁵。本稿では, *better off*が助動詞をはじめ, 法副詞,

²⁵ 査読者から, 第三ステージにおいて, 比較される状況の一つがなぜ仮想的な状況となるのかという質問をいただいた。この点については, *better off*の字義的な意味である「より豊かな, より幸せな」状態は, 一般的には望ましく, そうありたい状態であるため, 願望を表すような文脈で使われるうちに, *better off*と仮想的な状態の結びつきが強くなった可能性や, 助動詞 *had better*の使用頻度が1900年代頃に増えていることから, *better*の助動詞用法が広まったことに影響を受けた可能性等も考えられるが, 今後の課題とする。

if節などの、法に関する要素と共起する場合に、第三ステージに属すると判断をした。

- (6) (a) I believe we'd all be better off if everyone were to accept things philosophically.
[1947]
(b) Don't you think he'd be better off at the Benton Institute? [1935]
(c) You'd be much better off with me instead of somebody you didn't know. [1965]

(6)の文には、助動詞が含まれており、主節で表される事態が仮想的な状況であることを表している。また、主節後に仮想空間を表す *if* 節や前置詞句が、文の付加句として後続している。例えば、(6a)の「人々が物事を哲学的に受け入れる状況」、(6b)の「ある機関に所属する状況」、(6c)の「私という状況」は現実とは異なる仮想的な状況であり、その状況の中で主語はより良い状態にあると言える。一方で、(6a) (6b) (6c)における「人々が物事を哲学的に受け入れていない」「ある機関に所属していない」「私と一緒にいない」という現実の状況は背景化され言語化されていない。仮想世界における主語の状況が前景化されていることは、前節(4.3節)におけるコーパス調査の結果、*better off*と、モダリティを表す要素(助動詞、法副詞)や *if* 節との共起が増えたことから示唆される。また、仮想世界が前景化されると同時に、比較対象となる現状が背景化され、*better off*の比較の意味は薄れていく。このことは、*than*の出現頻度が継続して減っていることから示唆される。

第四ステージは、*better off*とそれに後続する要素に再分析が起こり、文の付加句(セッティング)として主語がより良い状態となる仮想的な状況を表していた要素が、主語の仮想的な状態を表すようになる段階である²⁶。この段階において、*better off*と後続する要素(=付加句)の結びつきが強くなり、[*better off* 補語句]として再分析されることで、*better off*構文が出現すると考えられる。補助記号を用いると再分析は(7b)と(7c)のように表せる。

- (7) (a) I'd be better off without you.
(b) I'd be [*better off*]_{complement} [*without you*]_{adjunct}
(c) I'd be [[*better off* [*without you*]_{complement}]_{complement}

(7b)は付加句(*adjunct*)の *without you*の構造を、(7c)は補語句(*complement*)の *without you*の構造を表したものである。(7b) (7c)は、文における付加句であった要素(*without you*)が、*better off*との結びつきが強くなり、最終的には *better off* XP全体が一つの構文としての地位を獲得したことを示している。ただし、*better off*に後続する要素が補語句であるか付加句であるかは排他的なものではなく、付加句の補語句としての再分析は段階的に進むものである。

この再分析を引き起こす要因として考えられるのが、文脈における語用論的な意味が、言語使用の中で、語彙の意味として定着することである(Hopper 1991)。つ

²⁶ 主語の状態を表すという点で主節補語と似た機能を持つと言える。

まり、(7b) の *You might be better off without her* を例とすると「彼女がい^{ない}状態が(いるよりも)より良い」という意味は、文脈内で「彼女がい^{ない}方がまし」な状況を表すことがあるが、同様の文脈で繰り返し用いられることにより、文脈から生じる「～の方がましだ」という語用論の意味は *better off* XP という句が持つ意味として再解釈される。この意味的な変化が、*better off* と後続する要素の結びつきを強め、文法面における再分析を引き起こしたと考えられる²⁷。

なお、前置詞句は付加句とも補語句ともみなすことができ解釈が難しいため、図1からは除外をしたが、本稿では、前置詞句の付加句から補語句への再分析によって *better off* 構文が誕生すると考えるため、前置詞に関する報告をここで行う²⁸。COHA 内に現れる *better off* 叙述用法の 2527 例の中で、前置詞句が後続するのは、639 例であった。年代別に集計をすると、*better off* に後続する要素が、前置詞句の場合(639 例)と前置詞句でない場合(1888 例)の割合はほとんど変わらなかった。一方、前置詞句内での各前置詞の割合には大きな変化が見られた。特に、前置詞句の 639 例の中で、仮想的な状況を表す *with* や *without* が、1900 年代以降、数が徐々に増えていることが確認できる。

表5 *better off* に後続する前置詞句と *with/without* 句の変化

年代	1820-1849	1850-1899	1900-1949	1950-1999	2000 ~	総計
<i>with/without</i>	2	29	82	141	39	293
他の前置詞	14	64	112	136	20	346
総計	16	93	194	277	59	639

以上の結果をまとめると、以下のような *better off* 構文の出現に関する仮説が提示できる。

- (8) *well off* の比較級 *better off* に意味の一般化と仮想的な状況の前景化が起こり「(仮想的な状況で)より良い」という意味を表す用法が強くなる中で、仮想的な状況を表す文の付加句が *better off* との結びつきを深め、*better off* の補語句として再分析されることで、「(仮想的な)状態の方がましだ」という意味

²⁷ 査読者より、*without* 以外で、付加句としても補語句としても解釈できるような例があると良いというご指摘をいただいた。付加句から補語句への再分析を起こす橋渡しとなるような文脈を含むと考えられる例を以下に記す。

(a) ... they would be better off with no legs, or with atrophied hands. [1974]

(b) She'll be better off in Al Shargaz, at least for a couple of weeks. [1986]

(a)(b) の *better off* に後続する前置詞句は、文の付加句としても解釈できるが、意味に注目すると、(a) の XP 句が表す「足が無い」あるいは「手が萎縮した」状態や、(b) の「(中東にあるとされる仮想の国である) Al Shargaz に居る」状態は一般的には好ましくない状態であり、全体で「～の方がましだ」という意味が生じているため、*better off* の補語句とも解釈できる。

²⁸ 本パラグラフは、査読者より COHA から抽出した *better off* の叙述用法の中で、*better off* に後続する要素が前置詞句である割合を示すことが重要との指摘を受けて、加筆をした。

を表す *better off* 構文が誕生した。

この再分析は, *better off* に助動詞との共起が増え, *than* との共起頻度が低くなった 1900 年頃にはすでに進行していたと考えられる。一方, この再分析の定着をコーパス内で明確な形として捉えられるのは, *better off* 後の補語句の中に, 不規則的な補語句が増え始める 1950 年頃である。すなわち再分析によって, *better off* に非時間的な要素を表す補語句が後続するようになった結果, 標準的な文法では不規則的とみなされる (9) のような, 形容詞, 分詞句が補語句の位置に現れる用法が増え始めたと考えられる。

(9) I'd be better off {dead, left alone, going home}

4.5 節では, 人手による *better off* の用例の分類と, 自然言語処理のアプローチを用いた用例のクラスタリングに基づくグループ化の比較を通して, (8) の仮説の妥当性を検討する。

4.5. 第四段階：仮説の検証と深層学習を用いた検証の補強²⁹

第四段階では, 第三段階で得られた, *better off* 構文が慣習化した過程に関する仮説に対して, 人手による *better off* の用例の分類と, 深層学習によるクラスタリングに基づくグループ化の経年変化の結果を比較しながら, 検証を行う。最初に, 人手による分類を行い仮説の妥当性を検証した後で, 深層学習に基づく言語処理的アプローチを用いることで検証結果を補強する。

深層学習に基づく検証結果の補強は以下のとおりである。第一に, *COHA* のノイズを除去した *CCOHA* の用例 (全 2591 例) を, 深層学習 (具体的には BERT) によりベクトル化する。第二に, 得られたベクトルに *k-means++* クラスタリング³⁰ を適用し自動的にグルーピングする。第三に, 各グループを年代ごとにまとめたグラフを作成することで用法変化を可視化する。この可視化した結果を, 人手の分類と比較することで, 第三段階で得られた仮説に一致するかどうかを吟味する。

本節での分析の基本は, 事例の分類結果を用法変化のグラフとして表すことにある。その際, グラフの横軸を年代, 縦軸を各分類グループの頻度とする。最初に, 人手による分類の方法を記す。ここでは, *better off* に起こった変化を捉えるため, 3 つのグループに分けて各用例を分類した。

表 6 人手による分類の内訳

①グループ 1：第一段階 + 第二段階：字義的な用法
②グループ 2：第三段階：仮想的な状況が前傾化した用法
③グループ 3：第四段階：確実に <i>better off</i> 構文とみなせる用法

²⁹ 本節の内容は永田・大谷・高村・川崎 (2022) に基づいている。

³⁰ <https://dl.acm.org/doi/10.5555/1283383.1283494> を参照

説明の都合上、グループ3から述べると、ここには、確実に *better off* 構文だと分類できる第四段階の用例（すなわち、補語句の位置に形容詞や分詞が現れるもの）が含まれる。さらに、*CCOHA*には、*better off*に *to* 不定詞が後続するもの (*better off to stay here*)があったが、これも、グループ3に分類をした。次に、グループ2は、仮想的な状況を想定した用法に対応する。具体的には、*would*や *may*などの助動詞、*if*節、*probably*、*maybe*などの法副詞を含む事例をグループ2とした。最後に、グループ3とグループ2の基準に当てはまらないような例をグループ1とした。なお、2591例に対する人手の分類の作業は、第一著者と第二著者で独立に行った。この時点での両者の一致率は95% ($k = 0.92$)であった。その後、すり合わせを行い、不一致の除去を行ったうえで、最終的なグループを決定した。

次に、深層学習を用いたグラフの作成手順を図示すると表7のようになる。

表7 深層学習を用いたグラフの作成手順

- | |
|--|
| ① <i>better off</i> の用例に対して BERT を用いてベクトルに変換 |
| ② 得られたベクトルを <i>k</i> -means クラスタリングでグループ化 |
| ③ 年代ごとにグループ内の用例数を求めグラフ化 |

以下に手順を具体的に記す。最初に、*CCOHA*から得られた2591の各事例を文単位でBERTに入力し、*better off*の出現位置に対応する最終層の隠れ状態を連結してベクトルとする。次に、得られたベクトルに *k*-means++ クラスタリング³¹ (Arthur and Vassilvitskii 2007)を適用し、グループ化を行う。グループの数については、4節で提案した、4つの段階に合わせたクラスタ数4と、表6の人手の分類に合わせたクラスタ数3のクラスタリングを行う。

図2と図3はそれぞれ、自動グループ化を用いて得た用法変化のグラフ（クラスタ数が3のもの）と、人手の分類を用いて得た用法変化のグラフである（永田・大谷・高村・川崎 2022）。

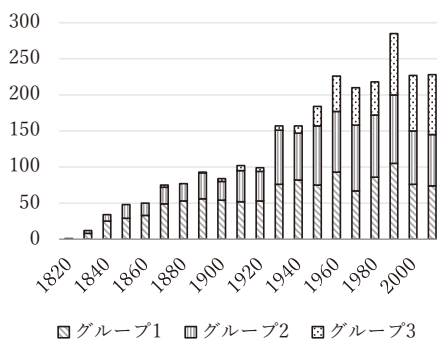


図2 自動グループ化を用いて得た用法変化

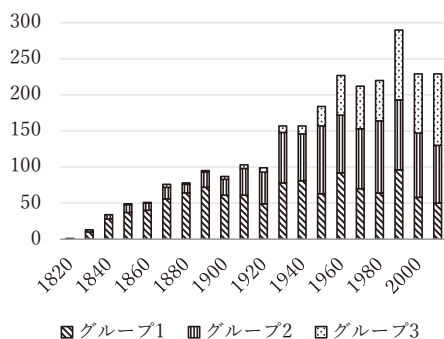


図3 人手の分類を用いて得た用法変化

³¹ 最初にデータを任意の数 (k 個) のクラスタ (グループ) に分けた後、データ間の類似度に基づいてデータを各クラスタに分かれるように調整させていくアルゴリズムの一種。

図2と図3では、縦軸が年代、横軸が各グループの事例数に対応する。二つのグラフ間の平均誤差は8.5であり、両図は非常に類似すると言える（人手の分類結果から得た図同士の平均誤差は2.1）。また、ラベルの一致率は0.72% ($k=0.56$)であった（人同士の一致率0.95 ($k=0.92$))。なお、図2は、クラスタリングを表すため、グループ番号の割り当てに任意性があるが、年代と頻度から決定した。すなわち、一番古い年代で頻度が最も高いグループを1、次に頻度が高いグループを2、残りを3とした。

まず、仮説の検証に先立ち、クラスタリングで得たグルーピングが人手の分類グループに対応するかを確認した。各グループの重心に最も近い事例10件を観察すると、グループ3については、全10件とも補語に現在分詞を取る事例であった（一方、グループ1と2については、一件も該当しなかった）。グループ3に分類された全事例を確認したところ、83% (=430/521) は、形容詞、現在分詞、*to* 不定詞のいずれかを従える用例であった。したがって、人の分類と同様に、グループ3は新用法（すなわち、*better off* 構文）に（ほぼ）対応すると言える。グループ2についても人の分類基準に対応することが確認できた。重心に最も近い事例の10件中9件は、*would* などの助動詞を伴う事例であった。残り1つは、*He believes they are better off here.* という、従属節の内容は主語が想定する事象であり、仮想的な状況を表すと解釈することができる事例であった。その他のグループ2の事例をランダムに36件サンプリングしたところ70%は人手の分類基準を満たす事例であった。グループ1は、完全には一致しないが、人間の分類基準に準ずることを確認した。重心に最も近い事例の10件のうちグループ3とグループ2の条件を満たすものはそれぞれ0件、3件であった。ただし、他の二つのグループに比較すると多様な事例を含んでおり、別の解釈も存在する可能性もある。以上の通り、人手の分類結果と、深層学習によるクラスタリングに基づくグループ化の経年変化の結果はおおむね一致しており、単語の出現分布から見ても *better off* の三つのグループは特徴的であることが示唆された。

次に、用法変化の図に基づいて *better off* に起こった変化をまとめると、図2、図3ともに、グループ3（確実に *better off* 構文とみなせる用例）が1960年代頃より増加し始めており、この時期以降は新用法が強く慣習化した第四ステージに入ったと言える。その少し前の1910～1950年代には、グループ2（仮想的な状況を表す語句を伴う用例）が増加傾向にある。ただし、人手による分類に基づいて作成した図3では、その増加はそれほど顕著ではないため、意味の一般化（第二ステージ）と並行して仮想的な意味の前景化（第三ステージ）が同時に起こったとも解釈できる。さらに、2000年代以降はグループ2が若干であるが減少している。これは、*better off* 構文が助動詞と共に起し仮想的な事態を表す用例の使用が増えた結果、全体の仮想的な意味が *better off* 構文の語彙の意味として定着したため、助動詞がなくとも仮想的な事態を表せるようになった可能性を示唆する。

以上の結果は、次のようにまとめられる。用法変化の図は、4.4節で提示した仮

説を大枠で支持する。特に、年代別の推移をみると、*better off*の叙述的用法に該当するグループ1から、仮想的な状況を表す語句をとまなうグループ2、*better off*構文に該当するグループ3が順に増えてきていることが明らかになった。さらに、注目すべきは人手による分類と機械によるグループ化が酷似している点である。BERTでは文脈情報を考慮してベクトル化した各用例をグループ化しており、各グループは文脈的な類似度により求められている。その3つのグループが1) 字義的なもの、2) 仮想的な状況における状態を表すもの、3) *better off*構文と確実に分類できるものに対応しており、また、各グループは1から3へと年代を経るにしたがって増えていった。このBERTに基づいた*better off*の用例のグループ化と、各グループの変遷は、(8)の仮説に基づいた人手による分類の結果ともかなりの程度一致したことから、BERTに基づいた*better off*の用例のグループ化の結果も、4.4節で挙げた仮説を支持する収束証拠の一つとみなせる。

本稿では、仮説に対して、人手による*better off*の用例の分類と、深層学習によるクラスタリングに基づくグループ化の経年変化の結果を比較しながら検証を行った。今回の研究では、人手によるコーディングの結果も仮説を支持しているため、人手の分析結果に加えて、わざわざ深層学習を検証に利用する必要はないという見方もできる。また、深層学習は文脈情報に基づいて用例をクラスタリングしてはいるが、実際にどのような基準に基づいて区別がなされているかについては精査が必要であるため、深層学習の結果が直接仮説を支持したわけではないという見方もできる。しかしながら、本稿の分析では、人手の分類と、深層学習に基づく用例のグループ化の結果がおおむね一致し、ともに仮説を裏づける結果を示したことから、人手だけの分類に比べて、認知言語学が提唱する収束証拠を深層学習の観点から提供している点で、言語学の伝統的な論証に比べ、客観性が高いと言える³²。

4.6. 言語研究サイクルの構築と自然言語処理の応用³³

本節では、前節までの研究を総括し、実証的な言語研究サイクルを具体的に提示すると同時に、言語研究サイクルの構築に自然言語処理が大きく貢献する可能性を記す。最初に、前節までに行った研究の四段階をまとめると以下のように提示できる。

³² 査読者から、本研究では、構文文法で仮定している変化プロセスが、深層学習によってダイレクトに証明されていないのではとの指摘をいただいた。本研究では、「言語学的な分類の妥当性を高める」という目的で、深層学習を用いたクラスタリングと人手の分析の一致度を見ているため、査読者が指摘するように、*better off*に生じたと考えられる変化プロセス（すなわち、再分析）に関しては検証をしていない。深層学習を用いた言語変化のプロセスの再現については興味深いテーマであるため、将来の課題とする。

³³ 本節の内容は大谷（2023）を発展させたものである。

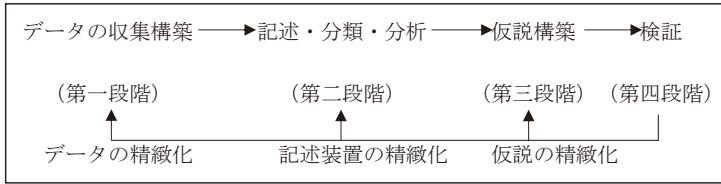


図4 実証的な言語分析のサイクルのイメージ

図4の言語分析サイクルにおいて、現在の構文文法や認知言語学がおもに従事しているのは、第一段階のデータの構築、第二段階の言語の記述・分類・分析、第三段階の仮説や理論の構築と言える。すなわち、構文文法や認知言語学では、フレームネット (Fillmore and Baker 2010) をはじめとしたデータの構築が行われており、また、コーパス、インフォーマント調査、文献、等から収集をしたデータに対して、妥当と思われる分類をしたり、分類をもとに、認知的に妥当と思われる仮説を提示したりする研究が盛んに行われている。一方で、仮説の検証については、言語学内部の手法で、ミニマルペアの比較やコーパスのデータの提示などが行われているものの、反証可能な形としての提示は、一部の学際的な研究 (Goldberg 2019 など) を除いてはなされておらず、追実験に至ってはほとんど行われていない状況にある。Dąbrowska and Divjak (2019) が指摘するように、現在、仮説を検証する方法論は、分野内でも分野外でも不足している状態にあると言える³⁴。一方で、言語研究が反証可能な科学的な研究になるためには、自然科学をはじめとした多くの分野で採用がされているように、仮説の立案と検証という分析のサイクルの中で研究を行うことが望ましい。

本節では、この研究サイクルに自然言語処理が大きく役立つ可能性を示す。最初に各段階における自然言語処理の活用方法を示す。第一に、データの構築の段階においては、特に、形式的な特徴のコーディングに関して、自動的なコーディングを用いることで、人手によるコーディングに比べ大幅に時間を短縮することが可能性となる。参考として記すと、本稿の4.4節で行われたコーディングは人手で行う場合は二週間程度かかったのに対して、BERTを用いたものは数十秒でコーディングが行われた。第二に、記述・分類・分析についても、本稿のグループ化したクラスタが示すように、少なくとも形式面に基づく集計や分類は人間が行うより短時間で行うことができる。第三に、仮説の構築についても、言語知識には明示的に捉えられる知識 (explicit knowledge) と潜在的な知識 (implicit knowledge) があるため、データ駆動型の研究を行うことが、潜在的な知識の発見に役立つ可能性が高い。例えば、本稿の4.3節の図1をみると、*better off* 構文が多く用いられ始めたのが1960年代以降であることが分かるが、20世紀以前もわずかながら出現していることも読み取

³⁴ Dąbrowska and Divjak (2019) は、今後の認知言語学の課題として、1) 実証的な研究手法の構築、2) 学際的な研究の展開、3) 社会的な転回 (social turn) の3点を挙げている。

れる。このような用例のレジスターや出現年を調べることによって、*better off*構文は特定のレジスターに最初に現れ、徐々に拡大したなどの仮説を生成することができる。第四に、仮説の検証に関しても、本稿で論じたように、用例の文脈をベクトル化したのちに、類似度に基づいてグループ化することで、多義の研究や意味変化の研究などに関する様々な仮説を実証する収束証拠の一つとして用いることができる。さらに、図4の矢印が示すように、仮説を実際に検証することで、各段階へのフィードバックも行うことができ、データの収集、記述・分類・分析、仮説構築の各段階を精緻化させることにつながる。

5. まとめと今後の課題：深層学習時代の構文研究に向けて

本稿では、*better off*構文がどのような過程を経て英語に定着したかを認知言語学、コーパス言語学、自然言語処理の観点から考察した。同時に、1) 現象の記述、2) コーパス調査、3) 仮説の生成、4) 仮説の検証という四段階に分けて論じることで、言語を実証的に分析するための研究サイクルの提示を行い、研究サイクルを構築するための各段階で自然言語処理の手法が貢献する可能性を論じた。さらに、近年発達した深層学習の理念を認知言語学の観点から解釈することで、深層学習の理念と用法基盤モデルには多くの類似点があることを指摘し、深層学習の技術が、認知言語学や構文文法の実証的な研究に役立つ可能性を示した。

最後に、これまでの節で得られた知見を基にして、深層学習時代の構文研究という観点から、言語研究における深層学習の位置づけ(5.1節)、深層学習を言語分析に用いる際の注意点(5.2節)と、その逆に、認知言語学や構文文法の知見が自然言語処理や深層学習に与える貢献の可能性(5.3節)について論じる。

5.1. 課題1: 言語研究における深層学習の位置づけ³⁵

本節では、深層学習を用いた言語分析の手法と、言語学の従来の方法論との比較を行いながら、言語学において深層学習がどのように位置づけられるかについて考察を行う。深層学習に基づいた言語モデルは、言語学における一般的な手法に比べ、はるかに多くの情報を文脈から抽出した分析を行う点で、言語学における様々な方法論の中でも、用法基盤モデルの理念に合致した方法論であると考えられる。一方で、深層学習に基づいた言語処理と人間の言語処理の関係性には不明な点も多く存在するため、現時点において、人間による分析にとって代わるような方法論とは言えない。したがって、本稿では、深層学習に基づいた言語モデルは、人間による研究を補強する手段の一つ(すなわち収束証拠)としての利用が現実的と考える。以下に、従来の言語分析を補強するような、深層学習の使い方を三点紹介する。

³⁵ 本節は、査読者から、人間による分析と深層学習に基づくクラスタリングの結果が一緒だったというだけでは、人間による分析と深層学習による分析の関連性が見えてこないため、両分析手法の異なりや優劣、その評価についての議論を追加した方が、有益な対比になるという意見を受け、加筆したものである。

第一に、本研究で用いたような、仮説の検証の補強としての使い方があり。人手による分析と、深層学習による分類の一致をみることで、実証的な裏付けが乏しく主観的な側面が含まれると批判されることがある、人間の分類の客観性が増すことが考えられる。言語学者の直観が言語学において果たしてきた役割は重要であるが、その直感が、文脈的な情報を最大限に考慮したクラスタリングと一致することによって、主観的と批判されることがある言語学的な分析の妥当性を高めることが可能となる。特に、今回用いた深層学習ベースの分類手法は、コーパスデータから単語列に内在する特徴を抽出し、分類基準を得るものである。言い換えれば、分類基準をあらかじめ与えたり、仮定したりはしない。そのようにして得られた分類結果が人間の分類結果と一致するという事は、人間の分類基準が作為的でなく、データに合致する自然なものであるという一つの根拠になる³⁶。

第二に、人手による言語データのコーディングには、長い時間がかかるが、人手の分類と深層学習の結果を組み合わせることで、短時間で効率よくデータを作成することが可能となる。つまり、人間によってある程度の量の事例のコーディングを行うと同時に、深層学習のクラスタリングを用いて、コーパスにおけるそれ以外の全事例を自動的に分類し、両者の結果の一致率が高い場合、人手による分類や仮説が、より広い範囲で成り立つ根拠として用いることができる。

第三に、深層学習をデータ駆動的に使用することで、仮説の生成にも役立つ。今回の分析では、先に仮説と言語的な分析があり、その追実験として深層学習を用いたが、逆に、前提を持たずに、データをクラスタリングしたり、あるいは、思い付き程度の考えに基づいて、データをクラスタリングすることで得られた結果に基づいて試験的に生成した仮説を、言語的な分析を用いて検証するという事も可能である。

このように、現在は、深層学習が言語研究に導入され始めた時期であるため、現時点で、深層学習の言語分析への応用は限定的かつ試験的なものととどまる。ただし、この点は、1960年代に生じたコーパス言語学の手法が言語学の主要な方法論として確立するまでに費やされた長い時間を考えると、当然と考えられる。深層学習が言語研究の方法論として言語学内部での地位を確立するまでには、深層学習に対する理解を深め、深層学習を用いた分析を積み重ね、言語学者が扱えるような便利なツールを開発するだけでなく、深層学習の結果が言語分析の結果と無関係なものとなさされないように、深層学習を言語理論に位置づけるという作業が決定的に重要であると考えられる。同時に、言語学者自身が深層学習の理論を学習したうえで、言語処理学者との共同研究を通じて、深層学習を用いた分析手法の提案をしていくことも重要と考えられる。

³⁶ ただし、今回用いた *k-means* クラスタリングでは、クラスタ数が3であるという事前知識を用いているので、厳密には分類基準を仮定しないとは言えない。

5.2. 課題 2: 深層学習を言語分析に用いる注意点

本稿ではこれまで、深層学習を用いた言語研究の例として *better off* 構文を論じ、深層学習が言語研究の様々な段階で役に立つとの主張を行ってきた。本稿では、用法基盤モデルは深層学習と極めて相性が良い言語理論という立場に立っている。その根拠となるのは、3節で論じたような、深層学習に基づく自然言語処理と用法基盤モデルの類似点である（表1を参照）。この類似点は驚くべきものであり、従来の機械学習に比べて、深層学習と用法基盤モデルの親和性が高くなっているとみなすことができる。

しかしながら、人間の言語習得と深層学習を比べた場合、本質的と思われる違いも見られるため、深層学習をそのまま人間の言語処理や言語知識の近似値とみなすのは危険がともなうという側面もある。本節では、深層学習を言語分析に用いる際に注意が必要な点を記す。

第一に、インプットされるデータの量である。人間が一生で受けるインプットに比べ、比較にならない量の言語データが、BERTにはインプットされる。第二に、インプットされる言語のタイプである。深層学習の場合、デジタル化しやすい書き言葉やオンラインのコンテンツが主なインプットである。一方、人間の言語習得におけるおもなインプットは話し言葉である。特に、言語習得の基盤とも考えられる、親が子供に向けて行う発話（Child-directed Speech）には、大人の会話には見られないような、繰り返しや大きな抑揚などの特徴が見られるため、書き言葉とはかなり質的に異なるものと言える。第三に、人間の言語のインプットは本質的に発話の場（いつ、だれが、どこで、どのような状況で）と切り離すことができないが、深層学習のデータにはそのような状況が付与されているとは限らない。第四に、人間はインプットを解釈して、その解釈に基づき行動をし、その結果、行動に対する相手の反応（表情や態度など）を通して、解釈の正誤がぼんやりした形ではあるが得られる。一方で、現時点での深層学習では、解釈に基づいた行動をしたり、その行動に対する相手の反応を通して、解釈の正誤を得たりすることはできない。第五に、人間のインプットには、言語情報に加え、パラ言語的情報や非言語的情報も多く含むが、実際の人間の言語使用に比べ、深層学習のインプットに含まれるパラ言語的情報や非言語的情報は限られる。

さらに、BERTをはじめとする深層学習に基づいた言語モデルは、人間の脳や言語処理過程を模倣して作られたというより、むしろ、翻訳などの言語処理の性能を向上させることを主な目的として発展してきたものであるため、言語モデルを、人間の言語処理のモデルに見立てて使用するには注意が必要である。同時に、言語モデル内で何が起きているかを明らかにするのは非常に困難なものであり、得られた結果の解釈は常に慎重に行う必要がある。

以上の点を含め様々な相違点があるため、深層学習が行う言語処理やその結果生じる言語知識が、人間が行う言語処理や人間が持つ言語知識とどの程度似たものであるかという点や、深層学習から得られた結果が実際に何を意味するのかという点

については、今後も検証を続ける必要がある。このような課題を踏まえても、本稿では、深層学習の結果を、認知科学における収束証拠の一つとして、人間の言語知識に関する様々な研究に適用することは効果的だと考える。さらに、言語学の分析に加えて、収束証拠として深層学習の結果を提示することは、言語学と自然言語処理の学際的な研究を促す効果もあると考える³⁷。

5.3. 課題3：構文文法から自然言語処理・深層学習への貢献

本稿では、これまで、自然言語処理のアプローチを用いた認知言語学の研究の紹介を行った。一方で、認知言語学や用法基盤モデルが行っている理論的な考察や研究によって蓄積された事実は、自然言語処理や深層学習へ貢献できる可能性も大いにある。例えば、現在でも、認知言語学におけるフレーム意味論の研究者が、言語処理の研究者と共同で構築しているフレームネットを用いて深層学習を再学習させる（=ファインチューニング）ことでモデルの精度が上がったという報告がなされている（山田・笹野・武田 2023）。また、深層学習を用いたロボットの記号創発において、二体のロボット間に共同注意（joint attention）を再現することで、記号創発が促進したとの報告もある（Taniguchi et al. 2016）。

本稿の最後に、深層学習というブラックボックスの解明に用法基盤モデルが生かせる可能性を指摘する。自然言語処理では、自然言語処理はそもそも工学であると同時に、Searle（1980）が示すように、機械による処理と実際の人間の脳による処理が同じかどうかを検証することは不可能との考え方もある。このような立場に立つ場合、深層学習というブラックボックスの中で実際に何が起きているかを検討すること自体そもそも不適切という結論に達するかもしれない。一方で、自然言語処理には、自然言語という人間の心から作り出された対象を扱うという点で認知科学の一員としての側面がある。このような立場から、ニューラルネット（例えば、大規模言語モデル）を人間の言語処理の働きの近似値とみなした場合、用法基盤モデルのような、言語使用から創発する言語知識を想定する言語学が、ニューラルネットの内部の解明や今後の発展に対して示唆する余地は大いにあるように思われる。

参考文献

- Arthur, David and Sergei Vassilvitskii (2007) k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms*, 1027–1035.
- Bergen, Benjamin K. and Nancy Chang (2005) Embodied construction grammar in simulation-based language understanding. *Construction grammars: Cognitive grounding and theoretical extensions*, 147–190. Amsterdam and Philadelphia: John Benjamins.
- Boas, Hans C. (2003) *A constructional approach to resultatives*. Stanford: CSLI Publications.
- Booij, Geert (2010) *Construction morphology*. Oxford: Oxford University Press.
- Bybee, Joan L. (1985) *Morphology: A study of the relation between meaning and form*. Amsterdam and

³⁷ 深層学習と認知言語学の学際的な研究の可能性については、大谷・永田・高村・川崎・持橋・大関・堀川（2024）で言語処理と脳科学の観点から試験的な検討を行った。

- Philadelphia: John Benjamins.
- Croft, William (2022) *Morphosyntax: Constructions of the world's languages*. Cambridge: Cambridge University Press.
- Cruttenden, Alan (1997) *Intonation* (2nd.). Cambridge: Cambridge University Press.
- Dąbrowska, Ewa and Dagmar Divjak (2019) *Cognitive linguistics: Foundations of language*. Berlin: Mouton De Gruyter.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2019) "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 4171–4186.
- Fillmore, Charles J., Paul Kay and Mary C. O' Connor (1988) Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64: 501–538.
- Fillmore, Charles J. and Collin F. Baker (2010) *A frames approach to semantic analysis*. In: Bernd Heine et al. (eds.) *Oxford handbook of linguistic analysis*, 313–341. Oxford: Oxford University Press.
- Goldberg, Adele E. (1995) *Constructions: A construction grammar approach to argument structure*. Chicago: The Chicago University Press.
- Goldberg, Adele E. (2006) *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, Adele E. (2019) *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Hasebe, Yoichiro (2023) Redefining the current discourse space model as a recursive monadic architecture. Oral at the 16th International Cognitive Linguistics Conference. HHU Düsseldorf, 8 August 2023.
- Hilpert, Martin (2019) *Construction grammar and its application to English* (2nd). Edinburgh: Edinburgh University Press.
- Hoffmann, Thomas (2022) *Construction grammar: The structure of English*. Cambridge: Cambridge University Press.
- Hopper, Paul J. (1991) On some principles of grammaticization. In: Elizabeth C. Traugott and Bernd Heine (eds.) *Approaches to grammaticalization*, 17–35. Amsterdam and Philadelphia: John Benjamins.
- Horikawa, Tomoyasu and Yukisaku Kamitani (2022) Attention modulates neural representation to render reconstructions according to subjective appearance. *Communications Biology* 5 (34).
- Iwata, Seizi (2008) *Locative alternation: A lexical-constructional approach*. Amsterdam and Philadelphia: John Benjamins.
- Janda, Laura A. (2013) *Cognitive linguistics: The quantitative turn*. Berlin and New York: Mouton de Gruyter.
- Kim, Jong-Bok and Laura A. Michaelis (2020) *Syntactic constructions in English*. Cambridge: Cambridge University Press.
- Lakoff, George (1990) The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* 1(1): 39–74.
- Langacker, Ronald W. (1987) *Foundations of cognitive grammar* (vol. 1): *Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. (1988) A Usage-based model. In: Brygida Rudzka-Ostyn (ed.) *Topics in cognitive linguistics*, 127–161. Amsterdam and Philadelphia: John Benjamins.
- Langacker, Ronald W. (2008) *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
- Mikolov, Tomas, Kai Chen, Greg Corrado and Jeffrey Dean (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- 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 回年次大会口頭発表, 2022 年 3 月 15 日.
- 中山俊秀・大谷直輝 (編) (2020) 『認知言語学と談話機能言語学の有機的接点—用法基盤モデルに基づく新展開』東京: ひつじ書房.
- 小野尚之 (2007) 『結果構文研究の新視点』東京: ひつじ書房.
- 大谷直輝 (2018) 「英語の *better off* 構文について」日本言語学会第 157 回大会口頭発表. 京都大学,

2018年11月17日.

- 大谷直輝 (2019) 『ベーシック英語構文文法』 東京：ひつじ書房.
- Otani, Naoki (2021) The rise of modal meaning: The case of *better off*. Oral at the 11th International Conference on Construction Grammar, University of Antwerp. 19 August 2021.
- 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回年次大会口頭発表, 2022年3月15日.
- 大谷直輝 (2023) 「言語知識の近似値として言語モデルは利用できるか」 『深層学習時代の計算言語学』 言語処理学会第29回年次大会ワークショップ, 2023年3月17日.
- 大谷直輝・永田亮・高村大也・川崎義史・持橋大地・大関洋平・堀川友慈 (2024) 「深層学習時代の言語研究」 『JELS』 221-225.
- Schönefeld, Doris (2011) *Converging evidence: Methodological and theoretical issues for linguistic research*. Amsterdam and Philadelphia: John Benjamins.
- Searle, John (1980) Minds, brains and programs. *Behavioral and Brain Sciences* 3 (3): 417-457.
- Steels, Luc (2013) How language emerges in situated embodied interactions. In: Lefebvre, Clare, Bernard Comrie and Henri Cohen (eds.) *New perspectives on the origins of language*, 505-532. Amsterdam and Philadelphia: John Benjamins.
- Steen, Francis and Mark B. Turner (2013) Multimodal construction grammar. In: Michael Borkent, Barbara Dancygier and Jennifer Hinnell (eds.) *Language and the Creative Mind*, 255-274. Stanford: CSLI Publications.
- Stefanowitsch, Anatol and Stefan Th. Gries (2003) Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209-243.
- 鈴木亮子・秦かおり・横森大輔 (編) (2017) 『話しことばへのアプローチ：創発的・学際的談話研究への新たな挑戦』 東京：ひつじ書房.
- Taniguchi, Tadahiro, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata and Hideki Asoh (2016) Symbol emergence in robotics: A survey. *Advanced Robotics* 30 (11-12): 706-728.
- Tomasello, Michael (ed.) (1998) *The new psychology of language: Cognitive and functional approaches to language structure* 1. Mahwah: Lawrence Erlbaum Associates.
- Tomasello, Michael (ed.) (2003a) *The new psychology of language: Cognitive and functional approaches to language structure* 2. Mahwah: Lawrence Erlbaum Associates.
- Tomasello, Michael (2003b) *Constructing a language: A usage-based theory of language acquisition*. Cambridge and Massachusetts: Harvard University Press.
- Torrent, Tiago T., Arthur L. Almeida and Mark Turner (2023) Co-Pilots for Frame Semanticists. Oral at the 16th International Cognitive Linguistics Conference. HHU Düsseldorf, 8 August 2023.
- 坪井栄治郎・早瀬尚子 (2020) 『認知文法と構文文法』 東京：開拓社.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems* 30 (NIPS 2017).
- Verhagen, Arie (2007) *Constructions of intersubjectivity: Discourse, syntax, and cognition*. Oxford: Oxford University Press.
- 山田康輔・笹野遼平・武田浩一 (2023) 「深層距離学習を用いた動詞の意味フレーム推定」 言語処理学会第28回年次大会口頭発表, 2023年3月16日.

コーパス

CCOHA: Clan Corpus of Historical American English
COHA: Corpus of Historical American English

執筆者連絡先：

e-mail: otani@tufs.ac.jp

[受領日 2023年9月4日

最終原稿受理日 2024年3月31日]

Abstract

**Investigating the Potential for Empirical Research on
Construction Grammar Using Deep Learning:
The Case of *Better Off* Construction**

NAOKI OTANI

Tokyo University of Foreign Studies

RYO NAGATA

Konan University

HIROYA TAKAMURA

*National Institute of Advanced
Industrial Science and Technology*

YOSHIFUMI KAWASAKI

The University of Tokyo

The present paper investigates how the so-called *better off* construction, an irregular idiomatic construction, emerged in the history of English from three perspectives: cognitive linguistics, corpus linguistics, and natural language processing (NLP). It then demonstrates various empirical methods to discuss constructions using deep learning. To that end, this paper (i) conducts extensive corpus research to examine the emergence of the *better off* construction, (ii) analyzes the data from a cognitive linguistic perspective to develop a hypothesis, and (iii) tests the legitimacy of the hypothesis using linguistic and NLP methods such as word vectors. Additionally, through the research on the *better off* construction, this paper attempts to (1) present a 'research cycle' for empirical linguistic analysis by applying various methods from NLP, (2) show the similarities between usage-based analysis and deep learning, and (3) investigate the possibility that deep learning can provide converging evidence for interdisciplinary linguistic analysis.